

## Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish<sup>☆</sup>

Pilar Prieto<sup>a,\*</sup>, Maria del Mar Vanrell<sup>b</sup>, Lluïsa Astruc<sup>c</sup>, Elinor Payne<sup>d</sup>, Brechtje Post<sup>e</sup>

<sup>a</sup> ICREA - Universitat Pompeu Fabra, Carrer Roc Boronat, 138, Despatx 51.600, 08018 Barcelona, Spain

<sup>b</sup> Universitat Pompeu Fabra, Carrer Roc Boronat, 138, Despatx 51.012, 08018 Barcelona, Spain

<sup>c</sup> Department of Languages, Faculty of Languages and Language Studies, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom

<sup>d</sup> Phonetics Lab, University of Oxford, 41 Wellington Square, Oxford OX1 2JF, United Kingdom

<sup>e</sup> Department of Theoretical and Applied Linguistics, Raised Faculty Building, Sidgwick Avenue, Cambridge CB3 4DA, UK

Received 29 December 2009; received in revised form 1 December 2011; accepted 2 December 2011

Available online 16 December 2011

### Abstract

The goal of this study is twofold: first, to examine in greater depth the claimed contribution of differences in syllable structure to measures of speech rhythm for three languages that are reported to belong to different rhythmic classes, namely, English, Spanish, and Catalan; and second, to investigate differences in the durational marking of prosodic heads and final edges of prosodic constituents between the three languages and test whether this distinction correlates in any way with the rhythmic distinctions. Data from a total of 24 speakers reading 720 utterances from these three languages show that differences in the rhythm metrics emerge even when syllable structure is controlled for in the experimental materials, at least between English on the one hand and Spanish/Catalan on the other, suggesting that important differences in durational patterns exist between these languages that cannot simply be attributed to differences in phonotactic properties. In particular, the vocalic variability measures nPVI-V,  $\Delta V$ , and VarcoV are shown to be robust tools for discrimination above and beyond such phonotactic properties. Further analyses of the data indicate that the rhythmic class distinctions under consideration finely correlate with differences in the way these languages instantiate two prosodic timing processes, namely, the durational marking of prosodic heads, and pre-final lengthening at prosodic boundaries.

© 2011 Elsevier B.V. All rights reserved.

**Keywords:** Rhythm; Rhythm index measures; Accentual lengthening; Final lengthening; Spanish language; Catalan language; English language

<sup>☆</sup> Earlier versions of this study were presented at the *Workshop on Empirical Approaches to Speech Rhythm* (28 March 2008, University College London, UK), at *Phonetics and Phonology in Iberia 2009* (17–19 June 2009, Las Palmas de Gran Canaria, Spain), and at *Speech Prosody 2010* (11–14 May 2010, Chicago, USA). We are grateful to the audience at these three conferences, and especially to M. D’Imperio, S. Frota, J.I. Hualde, F. Nolan, and L. White for fruitful discussions on some of the issues raised in this article. We also thank the action editor Jan van Santen and three anonymous reviewers for their comments on an earlier version of this paper. The idea for conducting the first experiment stems from an informal conversation with J.I. Hualde. We would like to thank N. Argemí, A. Barberà, M. Bell, A. Estrella, and F. Torres-Tamarit for recording the data in the three languages, and to N. Hilton and P. Roseano for carrying out the segmentation and coding of the data. Special thanks are due to F. Ramus et al. for making available to us the sentences used in their 1999 paper. This research has been funded by two Batista i Roca research projects entitled “The acquisition of rhythm in Catalan, Spanish and English” and “The acquisition of intonation in Catalan, Spanish and English” (Refs. 2007 PBR 29 and 2009 PBR 00018, respectively) awarded by the Generalitat de Catalunya, by the project “A cross-linguistic study of intonational development in young infants and children” awarded by the British Academy (Ref. SG-51777), and by the projects FFI2009-07648/FILO and CONSOLIDER-INGENIO 2010 “Bilingüismo y Neurociencia Cognitiva CSD2007-00012” awarded by the Spanish Ministerio de Ciencia e Innovación, and 2009 SGR 701, awarded by the Generalitat de Catalunya.

\* Corresponding author. Tel.: +34 93 542 22 48; fax: +34 93 542 16 17.

E-mail address: [pilar.prieto@upf.edu](mailto:pilar.prieto@upf.edu) (P. Prieto).

## 1. Introduction

As is well known, traditional studies on linguistic rhythm attempted to classify languages on a purely perceptual basis as being “syllable-timed” (e.g., Spanish and French) vs. “stress-timed” (e.g., English and Dutch); see Lloyd, 1940; Pike, 1945, and Abercrombie, 1967, among many others. This typological distinction was originally hypothesized to be instantiated through the isochronicity of speech intervals. That is, while syllables would tend to be of equal duration in syllable-timed languages, stress-delimited feet would tend to be of equal duration in stress-timed languages. This hypothesis has been called the isochrony hypothesis. However, instrumental studies have repeatedly failed to find acoustic evidence for constant or systematic duration in syllables or feet for either rhythm class. For example, several studies have shown that the duration of interstress intervals in English is not constant but rather varies in direct proportion to the number of syllables these intervals contain (Shen and Peterson, 1962; Bolinger, 1965; Roach, 1982; den Os, 1988, and others). Bolinger (1965) also showed that other factors such as the specific types of syllables as well as the position of the interval within the utterance have an effect on the duration of interstress intervals, Roach (1982) carried out measurements for the six languages analyzed by Abercrombie (1967), namely, French, Telugu, and Yoruba for the “syllable-timed” rhythm type, and Russian, English, and Arabic for the “stress-timed” rhythm type.<sup>1</sup> Against expectation, foot duration showed more variability in “stress-timed” languages than in “syllable-timed” languages. With respect to syllable duration, though the highest deviation was found in English and the lowest in Telugu (both in accordance with the hypothesis), the other four languages showed standard deviations of comparable magnitudes. As for “syllable-timed” languages, Borzone de Manrique and Signorini (1983) showed that syllable duration in Spanish is not constant and that interstress intervals tend to cluster around an average duration. Similarly, Wenk and Wiolland (1982) did not report isochronicity in syllable duration in French. Instead, they proposed that larger rhythmic units of the size roughly corresponding to the phonological phrase in prosodic phonology are responsible for rhythm in French.

The lack of clear acoustic evidence for distinct isochronous units in speech led to the development of an alternative view of linguistic rhythm, which has often been called the *phonological approach to language rhythm* (see Ramus et al., 1999).<sup>2</sup> According to this view, the percept of

different types of rhythm is the result of the presence/absence of specific phonological and phonetic properties in a particular language system. Dauer (1983) observed that stress-timed and syllable-timed languages typically correlate with a number of different distinctive phonetic and phonological properties such as syllable structure variety and complexity, vowel reduction, and differences in the correlates of stress. She claimed that the coexistence of a certain set of these phonological properties is responsible for promoting the perceptual prominence of stressed syllables in relation to other syllables – yielding “stress-timed” perception – while a different set is responsible for the percept of equal salience between syllables – yielding “syllable-timed” perception. Most of this work on linguistic rhythm assumes that the differences in rhythmic percept found between languages arise from distinct combinations of these phonetic and phonological properties, of which the two most cited are syllable structure and vowel reduction. The so-called “stress-timed” languages, like English, have a greater range of syllable structure types, allowing for more complex codas and onsets, and also tend to reduce unstressed vowels both durationally and qualitatively (Delattre, 1966). By contrast, the so-called “syllable-timed” languages, like Spanish, have a significant amount of open syllables and no vowel reduction. According to Dauer (1987), a language cannot be assigned to one or the other rhythmic class on the basis of instrumental measurements of interstress intervals or syllable durations.

Ramus et al. (1999) pioneering article set out to study the correlates of linguistic rhythm that could be found in the phonetic stream, since they argued that a viable account of speech rhythm could not rely exclusively on complex and language-dependent phonological concepts. They found that three acoustic indices based on vocalic and consonantal interval measures, namely, %*V* (or amount of vocalic stretch per utterance),  $\Delta V$  and  $\Delta C$  (standard deviations in the duration of the vocalic and consonantal stretches in the utterance respectively) enabled newborn infants to distinguish between perceived rhythmic classes. Ramus et al. extended these insights of the phonological account of rhythm. They pointed out that “even though the phonological account appears to be adequate, it does not explain how rhythm is extracted from the speech signal by the perceptual system”. (p. 269). As Ramus et al. (1999, p. 265) conclude, “the measurements suggest that intuitive rhythm types reflect specific phonological properties, which in turn are signalled by the acoustic/phonetic properties of speech”. Parallel work on speech rhythm has developed a range of metrics that have proved able, with varying degrees of success, to discriminate and capture the differences between language rhythmic classes. One of these is the Pairwise Variability Index, or PVI (Low et al., 2000; Grabe and Low, 2002; Asu and Nolan, 2006). In addition, rate-normalised versions of Ramus’  $\Delta C$  and  $\Delta V$  measures have been proposed (VarcoC and VarcoV, respectively, see Dellwo, 2006, and Ferragne and

<sup>1</sup> Even though we acknowledge that the motivation for the terms “syllable-timed” and “stress-timed” has been discredited by empirical studies, we will use them as a convenient short-hand for referring to two categories that have been shown to be perceptually distinct in rhythm terms.

<sup>2</sup> Ramus et al. (1999) were the first to refer to Dauer’s approach with this term: “Henceforth we will refer to this position [Dauer’s position] as the phonological account of rhythm” (p. 269).

Pellegrino, 2004). These measures are reviewed and compared for discriminatory effectiveness in (White and Mattys, 2007a). See also Section 1.5 in this paper for detailed descriptions of these metrics.

Following Ramus et al. (1999), investigations on linguistic rhythm have commonly taken the view that it is the phonological makeup of sentences that forms the basis of perceived rhythmic differences between languages. The materials used by these studies are argued to typically reflect the language-specific phonological properties of the language being investigated (e.g., Carter, 2005; O'Rourke, 2008; Nolan and Asu, 2009 for Spanish, Frota and Vigário, 2001 for European and Brazilian Portuguese, Asu and Nolan, 2006 for Estonian, White and Mattys, 2007a,b for Dutch, French, English, and Spanish, and Russo and Barry, 2008 for Italian and Dutch, among others). Among the phonological properties that are related to rhythm, syllable structure is one of the most frequently and reliably cited.<sup>3</sup>

However, it is also well known that higher levels of prosodic structure such as prominence marking and prosodic phrasing strongly influence the organization of timing across languages. Crosslinguistic evidence demonstrates that increased duration is an important acoustic correlate of prosodic heads (or prominent units) and edges of prosodic constituents. For example, it has been shown that stressed and accented syllables are produced with additional lengthening compared with unstressed syllables (e.g., Beckman and Edwards, 1994; Turk and Sawusch, 1997, and Turk and White, 1999, among many others). Similarly, the edges of prosodic constituents have been shown to trigger lengthening effects crosslinguistically. Word-initial lengthening has been reported by Oller (1973), Dilley et al. (1996), Fougeron and Keating (1996), Byrd (2000), and others; and phrase-final lengthening has been reported by Wightman et al. (1992) and Yoon et al. (2007) among many others. Final lengthening at the edges of phrasal prosodic constituents is a very widespread, possibly universal, phenomenon across languages (see Beckman, 1992). However, while phenomena like initial strengthening and final lengthening may potentially be universal, they have also been shown to be subject to language-specific variation in their phonetic implementation. For example, Barnes (2001) claims that the durational asymmetry found between Turkish and English vowels in initial and non-initial syllables can be attributed to language-specific instantiations of initial strengthening. Similarly, the durational marking of prominent syllables (stressed and accented syllables) has been reported to vary

between languages (Beckman, 1992; Ortega-Llebaria and Prieto, 2007; Hualde, 2005; Fant et al., 1991a).

Recently, several authors have pointed to the need for an examination of phrasal timing phenomena across languages in relation to the percept of rhythmic classes. As Beckman (1992) states, “it should be useful to compare timing patterns across languages in order to relate whatever similarities and differences we find to the universal or language-specific aspects of linguistic structure. (...) We need to go inside the larger unit, as Fant, Kruckenberg, and Nord so nicely put it, and examine the actual lengths of component syllables as a function of their position within the hierarchy of stresses and phrases”. (Beckman, 1992, p. 457). Similarly, White and Mattys (2007a, p. 518) point out that “language-specific prosodic timing processes, such as accentual lengthening, word-initial lengthening, and phrase-final lengthening, should clearly also be considered in a full model of influences on vocalic and consonantal interval durations. How far rhythm class distinctions correlate with differences in prosodic timing processes is an open question”. White et al. (2009) re-emphasize the need to take prosodic timing processes into account, citing evidence of critical differences in these processes that appear to correlate with perceived rhythmic differences, in their comparison of northern and southern varieties of Italian (see also Arvaniti, 2009).

The overarching goal of this paper is to explore the role of syllable structure and other prosodic timing processes (like phrasal prosodic processes) in the instantiation of durational and rhythmic differences across three well-investigated languages. We will compare three languages that traditionally have been claimed to lie at different points on a hypothesised rhythmic scale, namely, Spanish (a prototypical “syllable-timed” language), English (a prototypical “stress-timed” language), and Catalan (claimed to be a mixed or “intermediate language” in the rhythmic scale). “Intermediate” languages like Catalan or Polish have been posited because they exhibit phonological properties that are associated with both types of rhythm class (Nespor, 1990). For example, even though Catalan has been described as a “syllable-timed” language in terms of properties like phonotactics, it also has vowel reduction, which is typically associated with “stress-timed” languages. Similarly, Polish has many properties that would classify it as a “stress-timed” language, but it does not exhibit vowel reduction (Nespor, 1990). Therefore, having some features distinctive of “stress-timed” languages and others distinctive of “syllable-timed languages”, Catalan would be predicted to behave rhythmically as an intermediate language. Yet the acoustic evidence in this respect is somewhat contradictory. While Ramus et al. (1999) find that Catalan clusters with other “syllable-timed” languages such as Spanish, Italian, and French, Grabe and Low (2002) report that Catalan vocalic pairwise variability measures differ from those of Spanish, with an intermediate rhythm index. Gavalda-Ferré (2007) undertook a study that compared Catalan with other languages included in the BonnTempo

<sup>3</sup> As Ramus et al. (1999, p. 289) point out, “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 (...). Given the current state of knowledge, we believe that only syllable structure is reliably related to rhythm”.

corpus (Dellwo et al., 2004). In accordance with the previous investigations, Gavaldà-Ferré finds that data for %*V* and  $\Delta C$  classify Catalan as a “syllable-timed” language whereas results for nPVI and rPVI characterize Catalan as an intermediate-timed language. The analysis of adult-directed speech carried out by Payne et al. (in press) showed that Catalan is distinct from English in all of the scores but is distinct from Spanish only in the normalized vocalic variability metrics (VarcoV and nPVI-V). In summary, most of the evidence gathered thus far suggests that Catalan tends to pattern more closely with Spanish than with English for most rhythm metrics.

In order to investigate the potential contribution of these two language-specific properties (syllable structure complexity and prosodic timing phenomena) to rhythmic measures, the empirical materials used in this study were controlled for syllable structure composition and prosodic structure grouping. The experimental materials consisted of the following two types of utterances: (a) the same English, Spanish, and Catalan utterances used by Ramus et al. (1999), which were not controlled for syllable structure types; and (b) two sets of utterances controlled for syllable composition, namely, utterances mostly composed of open CV syllables and utterances composed of closed CVC syllables, all of them having similar phrasal prosodic properties. We recorded 24 female speakers per language reading a total of 30 utterances (for a total of 720 utterances, 30 utterances  $\times$  8 speakers  $\times$  3 languages).

With regard to the first specific goal of the study, if differences in syllable structure complexity are a crucial factor leading to perceived differences in rhythm, we would expect to find non-significant differences in the metrics between the two sets of controlled materials (the CV materials and the CVC materials). If, on the contrary, differences in the rhythm metrics remain when controlling for syllable type, this would suggest that speech rhythm does not emerge exclusively from language-specific phonotactic properties, and that other processes that affect timing are at work (see also Arvaniti, 2009). Our intuition would lead us to expect the latter, since a word such as *banana* sounds rhythmically different when spoken by a Spanish or an English native speaker, regardless of the fact that this word is composed exclusively of open CV syllables in both languages.

The second goal of the study is to examine whether the perceived differences in rhythm across the three languages are correlated with prosodic timing differences, namely, the durational marking of prosodic heads (i.e., accentual lengthening) and prosodic edges (i.e., final lengthening). Our initial hypothesis is that prosodic timing differences will correlate with rhythmic differences. English, being a “stress-timed” language, is known to show a lengthening effect both in stressed syllables and at the edges of prosodic domains, and thus we expect higher vocalic interval scores and vocalic variability scores than in Spanish across the two types of controlled materials (open vs. closed syllable). This would provide compelling evidence in support of the

hypothesis that language-specific phrasing timing phenomena play an important role in the creation of the percept of rhythmic differences.

This article is organized as follows. First, the second section describes the language corpus materials used for the production experiment and the methodology used for the prosodic labeling of the data. Second, the third section is divided into two parts. In the first part we present the main findings on the effects of syllable structure on the different rhythm metrics. In the second part we analyze the behaviour of two prosodic timing processes in the data, namely, lengthening patterns related to prominence-lending units and prosodic boundaries. Finally, the fourth and fifth sections evaluate the implications of the results for our understanding of linguistic rhythm.

## 2. Methodology

### 2.1. Target languages

The three languages chosen have consistently been documented as prototypical examples of “stress-timed” (English), “syllable-timed” (Spanish), and “intermediate-timed” languages (Catalan). As detailed below, the three languages display a variety of different phonological and prosodic properties. Importantly, Catalan has a mixed type of behavior that will be of particular interest for our purposes:

- (i) **English** has been traditionally classed as a “stress-timed” language. It displays the typical phonological properties of this class, namely, a wide variety of syllable structure types, a high frequency of complex syllables structures, vowel reduction in unstressed positions, and substantial final lengthening (Wightman et al., 1992).
- (ii) **Spanish** has been traditionally classed as a “syllable-timed” language. The predominant syllable type is CV and it shows low degrees of syllable complexity. In addition, it displays almost no signs of vowel reduction and less final lengthening than English (Hualde, 2005; Ortega-Llebaria and Prieto, 2007).
- (iii) **Catalan** has been classed as an “intermediate” language between “syllable-timed” and “stress-timed” because of its mixed phonological properties (Nespor, 1990). Though Catalan shows greater complexity in terms of syllable structure types than Spanish and also in terms of the proportion of closed syllables than Spanish (e.g., Span. *caballo*, Cat. *cavall* ‘horse’; Span. *arco*, Cat. *arc* ‘arch’), the predominant syllable type is CV. It also presents vowel reduction, a property which has been consistently associated with “stress-timed” languages.

As we have noted, though Catalan has been classified as an intermediate language, phonologists have not reached complete agreement on its rhythmic status. Nespor (1990) was the first to propose Catalan as an intermediate



language. She noticed that Catalan “has 12 most common syllable types that are constituted by a minimum of 1 and a maximum of 6 segments” (Nespor, 1990, p. 164). This would set Catalan apart from the prototypical “syllable-timed” languages like Italian, Greek, or Spanish, which have a more limited range of syllable structure types, and bring it closer to “stress-timed” languages such as Dutch and English, which have a greater range of syllable structure types. In addition, Catalan displays a process of vowel reduction, by which the system of seven vowels that may occur in stressed positions /i/, /e/, /ɛ/, /a/, /o/, /ə/, and schwa are reduced to only three /i/, /u/, and schwa in unstressed positions (see Mascaró, 2002 for a thorough review of this process). Thus, when a syllable loses its stress, /e/, /ɛ/, /a/ are reduced to schwa, /o/ and /ɔ/ are reduced to /u/, and only /i/ and /u/ retain their vowel quality. Finally, Catalan, like Spanish, shows much less utterance-final lengthening than English (Ortega-Llebaria and Prieto, 2007).

## 2.2. Materials

The experimental materials used in this investigation fall into three main types. First, we designed a set of controlled materials which consisted of 10 utterances per language which were matched for utterance length (number of syllables and number of prosodic/orthographic words)<sup>4</sup> and syllabic structure. Half of these utterances were composed of predominantly open CV-type syllables (CV-type sentences) and the other half of predominantly closed syllables (CVC-type sentences, mainly CVC and occasionally CVCC). All of these utterances were fairly well matched for number of syllables (from 13 to 19) and for segmental and prosodic word composition. The percentage of closed syllables in the “Closed Syllable condition” was 95.3% for Catalan, 86.7% for English, and 90.2% for Spanish. The percentage of open syllables in the “open syllable condition” was 89.4% for Catalan, 87.2% for English, and 89.1% for Spanish. By contrast, the proportion of open syllables in the “Mixed condition” (see below) was 66.5% for Catalan, 45.6% for English, and 70.2% for Spanish. Further, these materials contained target lexical items which are homophonous in the three languages. For this purpose we exploited the use of place names (Guatemala, Barcelona, Brazil, Ceylon, etc.), as well as cognates and borrowings (e.g., “international”, “mangos”, “tennis”, “meeting”, “oranges”, or “parking”). Importantly, given that the number of prosodic words was controlled for, the expectation was that target utterances would be pronounced in two prosodic phrases (see Frota et al., 2007 for the patterns of phrasing

found across Romance language SVO sentences). In our materials, the mean number of prosodic words ranged between 5 and 6 (5.4 for Catalan, 6.2 for English, and 5.1 for Spanish), and these were expected to be pronounced in two (sometimes three) prosodic groups (see results below). As for number of orthographic words, the materials ranged from a mean of 8.7 for Spanish to 10.4 for English (with 9.50 for Spanish).<sup>5</sup>

Second, we used a set of “mixed materials” that were representative of the particular target language and that would act as control sentences in the analysis of the effects of syllable structure.<sup>6</sup> These mixed materials consisted of the set of sentences used by Ramus et al. (1999) for Catalan, English, and Spanish. These sentences are short news-like declarative statements that are matched across the three languages for number of syllables (from 14 to 19).

In sum, the materials used for this study belong to one of three categories, namely, CV-type sentences (or sentences with predominantly open or CV syllables), CVC-type sentences (or sentences with predominantly closed CVC syllables), and mixed-type sentences (or sentences which have an uncontrolled number of open and closed syllables). The examples in (1) show one target utterance from each language, for each of the categories (CV-type, CVC-type, and mixed), with the number of syllables per sentence appears in parentheses. The whole set of target utterances can be found in the Appendix A.

- 
- (1) CV-type utterances (predominantly open syllables)  
 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)
- CVC-type utterances (predominantly closed syllables)  
 Cat: Els dònuts d’Amsterdam són realment internacionals (15)  
 Eng: These doughnuts from Amsterdam taste almost exceptional (14)  
 Span: Los donuts de Amsterdam son realmente internacionales (15)
- Mixed (utterances taken from Ramus et al., 1999)  
 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)
- 

<sup>4</sup> Prosodic words are typically characterized as being the domain of word stress, phonotactic constraints and phonological word-level rules. In our materials we took primary stress to be the main predictor of the presence of a prosodic word. For example, the Spanish sentence *Se enteraron de la noticia en este diario* has 8 morphological words and 4 prosodic words.

<sup>5</sup> Results from three Wilcoxon matched pairs signed rank tests for number of prosodic words and number of orthographic words for each sentence as dependent variables revealed non significant differences between groups except for the comparison between Spanish and English ( $T = 16, p < .0001, r = -.72$  and  $T = 25, p < .01, r = -.60$  for the number of prosodic words and orthographic words, respectively). Significant differences were also found between Catalan and Spanish only for the variable number of prosodic words ( $T = 20, p < .01, r = -.55$ ).

<sup>6</sup> We would like to thank Ramus et al. for sending us the Spanish, English, and Catalan sentences used for their 1999 paper.

### 2.3. Subjects and recording procedure

The 30 target utterances were read at a normal speech rate by a total of 24 speakers: 8 Southern Standard British English speakers from the Cambridge area, 8 Central Peninsular Spanish speakers from the Madrid area, and 8 Central Catalan speakers from the Barcelona area. All participants in this study were females between the ages of 28 and 40.

Recordings were made in a quiet room in the participants' homes in Cambridge, Madrid and Barcelona, respectively, using a Marantz PMD660 recorder and Shure PG81 microphones for the Spanish and Catalan recordings, and a Tascam HD-P2 recorder with AKG C3000B microphones for the English recordings. Subjects were given time prior to the recordings to practice reading the sentences. When errors or hesitations occurred during the recording, subjects were asked to repeat the tokens at the end of the session. This reading task was performed as a part of a long recording session, which included other tasks intended to investigate the rhythmic and intonational properties of these languages in children's and child-directed speech (see *Astruc et al., 2009; Payne et al., 2009, 2010, in press, Vanrell et al., 2010*).

We first analyzed the phrasing patterns in the data to check whether the prosodic grouping of sentences was close to what we would expect. In the CV condition, the average number of prosodic groups was very close to 2 in the three languages (between 2.1 in English and Spanish and 2.3 in Catalan). In the CVC condition, the number of prosodic groups was slightly higher, ranging from an average of 2.5 in both Catalan and English to 2.6 in Spanish. In the mixed materials the average number of prosodic words ranged from 2.3 in Catalan and Spanish to 2.8 in English. Also, to check whether speech rate was comparable across language groups, we calculated the speech rate in number of syllables per second as a function of the phonotactic materials. The results revealed that though the English materials were produced with a slightly lower speech rate (a mean of 5.3 syllables/s) than Catalan and Spanish (with a mean of 5.9 and 6.1 syllables/s respectively), the variation was quite high across phonotactic conditions. An univariate ANOVA was run to the speech rate data (in syllables per second), with Language (3 levels: Catalan, Spanish, English), Syllable Type (3 levels: open, closed, mixed) and Number of Syllables (from 14 to 19) as fixed factors. The three factors were significant at  $p < 0.001$ . The paired post hoc analyses revealed a significant difference between the three pairs of languages (at  $p < 0.001$ ), namely between Catalan and Spanish, Catalan and English, and Spanish and English. Crucially, the partial eta squared estimates reveal that the only moderate magnitude effect on speech rate comes from the factor Number of Syllables ( $\eta^2 = 0.269$ ). By contrast, the estimate effect of the two other factors, namely Language ( $\eta^2 = 0.113$ ) and Syllable Type ( $\eta^2 = 0.135$ ), is medium low. This means that the estimate effect of Language is as small as Syllable Type in the

prediction of speech rate. Focusing also on the potential speaker effects, we decided to run separate analyses for each language, but now having Speaker as a fixed factor rather than Language. The same results were obtained, and all factors including Speaker were significant at  $p < 0.001$  for each Language, revealing large speaker differences within a language.

The total number of utterances analyzed was 720 (8 speakers  $\times$  30 utterances  $\times$  3 languages). The total number of syllables analyzed was 12,086, and the total number of segments analyzed was 29,151. All the data is available from the project website: <http://www.april-project.info/>.

### 2.4. Data segmentation

The main segmental and prosodic labeling was performed by a research assistant, a trained phonetician who is a native speaker of English and also proficient in Spanish.<sup>7</sup> The Catalan and Spanish labeling was checked and where necessary revised by the three author native speakers of these two languages. Vocalic and consonantal segmentation was performed using Praat (*Boersma and Weenink, 2007*) and according to standard criteria, e.g., F2-onset for vowels (*Peterson and Lehiste, 1960*). For instance, in fricative-vowel sequences, the onset of the vowel was taken to be the beginning of the second formant. The onset of a fricative was marked at the start of high frequency energy. Prevoallic glides were considered to be part of the consonantal intervals and postvoallic glides part of the vocalic intervals (e.g., the first syllable of *Guatemala* was treated as CCV; the first syllable of *Ceílán* as CVV). Syllables were separated according to Catalan and Spanish syllabification rules, by which CV structures are maintained whenever possible and there is a CV resyllabification process across word boundaries. For English, as is well known, syllable boundaries are more difficult to determine: segments were placed in onsets in preference to codas, except when the acoustics indicated that a segment belonged to the coda of the preceding syllable (i.e., consonants with lower intensity or shorter duration).

Prosodic labeling was also performed on the language materials. We marked those prosodic phenomena which have been claimed to have a particularly strong influence on timing patterns, namely, proximity to prosodic boundaries and differences in prominence level. With respect to prosodic phrasing, the three languages were labeled with two levels of phrasing, namely, the intonational phrase (or *IP*) and the intermediate phrase (or *ip*; see the original *Beckman and Pierrehumbert, 1986* proposal for English). The latter intonationally-based constituent, whose boundary corresponds to a level 3 Break Index in the ToBI system, is defined as an intonation contour with one or more pitch accents and a final phrase accent. This phrasing

<sup>7</sup> We would like to thank Naomi Hilton for performing the segmental and prosodic coding of the data in the three languages. Thanks also go to Paolo Roseano for revising the prosodic labeling.

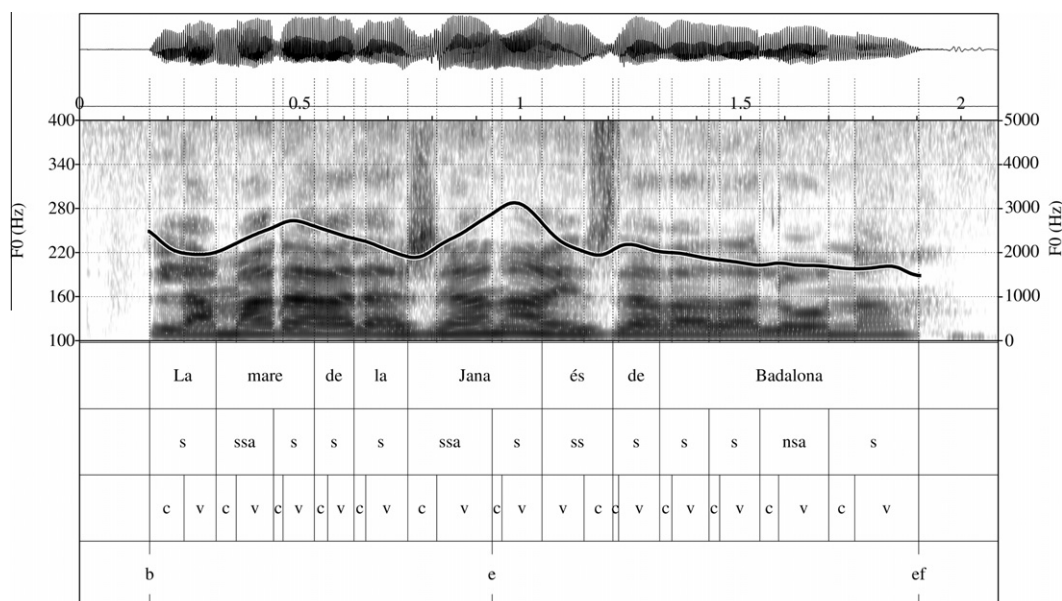


Fig. 1. Waveform, spectrogram, f0 contour, and labeling schema used for the Catalan utterance *La mare de la Jana és de Badalona* ‘Jana’s mother is from Badalona’ (speaker mSMN).

distinction has also been proposed for the Catalan and Spanish versions of ToBI, *Cat\_ToBI* (Prieto, in press) and *Sp\_ToBI* (Beckman et al., 2002; Estebas-Vilaplana and Prieto, 2010).<sup>8</sup> In addition to tonal marking, which must be present at the end of both *IPs* and *ips*, durational lengthening has also been claimed to be of a lesser intensity at the end of the former constituent (Wightman et al., 1992; Yoon et al. (2007) for English). For the purposes of our labeling, the criterion for an *IP* break was the presence of a pause of at least 200 ms. In our data, a good portion of the utterances were produced in two intermediate phrases, generally produced with a continuation rise located at the end of the first intermediate phrase.

With respect to phrase-level prominence categories, following studies on English prominence, e.g., Beckman and Edwards (1994), we assume that stress is used to convey prominence at all levels of the prominence hierarchy, and that this prominence is cumulative across levels. In our data, we distinguished the following: unstressed, lexically stressed, accented, and nuclear accented syllables, which have all been associated with significant lengthening (for accentual lengthening, see for example Turk and Sawusch, 1997; Turk and White, 1999).

Fig. 1 illustrates the orthographic, segmental, and prosodic transcription of the Catalan utterance *La mare de la Jana és de Badalona* ‘Jana’s mother is from Badalona’. 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, following the standard procedures explained above. Finally, the fourth tier contains the phrasing information: beginning of a prosodic domain (=b), end of an intermediate phrase (=e), and end of an intonational phrase (=ef), together with pause markings (=p).

An inter-transcriber reliability test for the prosodic coding (accentuation and phrasing) was conducted with a subset of our data. Ten percent of the data (a total of 144 utterances) from the database were randomly selected by two of the authors, taking into account that they were uniformly represented across languages, speakers, and conditions (mixed, open-syllable, and closed-syllable conditions). Five transcribers (the authors of the paper) then labeled the target utterances independently using the *Cat\_ToBI*, *Sp\_ToBI*, and *Eng\_ToBI* systems respectively, after having a couple of sessions of hands-on discussion on a set of examples. The guidelines followed by the labelers were the same as the descriptions of accentuation and phrasing provided in this section. The labelers had to decide on three levels of accentuation (unaccented, accented, and nuclear accented), and two levels of phrasing (intermediate phrase, intonational phrase).

A comparison of the prosodic transcription across the five transcribers revealed a very high consistency rate both in accentuation and phrasing decisions. Overall agreement on the choice of phrasing level (three levels: no break, intermediate phrase, intonational phrase) was very high: 97.7% consistency for Catalan, 97.9% consistency for English, and 98.5% consistency for Spanish. Likewise, agreement on the choice of prominence levels (three levels: unaccented, accented, nuclear accented) was 94.2% for Catalan,

<sup>8</sup> The reader can access the *Cat\_ToBI Training Materials* and the *Sp\_ToBI Training Materials*, together with audio files and exercises, at the following web pages: [http://www.prosodia.upf.edu/cat\\_tobi/en/index.php](http://www.prosodia.upf.edu/cat_tobi/en/index.php) (*Cat\_ToBI*) [http://www.prosodia.upf.edu/sp\\_tobi/en/index.php](http://www.prosodia.upf.edu/sp_tobi/en/index.php) (*Sp\_ToBI*).



92.8% for English, and 95.7% for Spanish. Our agreement rates are similar to previous rates from other ToBI studies (Syrdal and McGory, 2000; Yoon et al., 2004). The relatively high agreement rates on the choice of phrasing level (three levels: no break, intermediate phrase, intonational phrase, between 97% and 98%) and prominence level (three levels: unaccented, accented, nuclear accented, between 94% and 95%) are almost certainly due to the smaller set of categories in the labeling that our transcribers had to perform. For example, our results are similar to some of the previous ToBI results on the presence or absence of a pitch accent (91.5% of agreement in the case of Syrdal and McGory's (2000) study and 89.14% in Yoon et al.'s study (2004)). This analysis is similar to the prominence level choice in our study, with only one more category involved in our case, namely, the nuclear accent choice. With respect to the phrasing transcription, our study also reveals similar agreement results to previous results on edge tone choice (92% of agreement in the case of Syrdal and McGory's (2000) study and 88.39% in Yoon et al.'s study (2004)). Taking into account that phrasing transcription of our data was a simple ternary decision on the phrasing level of the boundary (i.e., absence of boundary, and intermediate phrase or intonational phrase in the case of presence of a boundary), such a small number of categories can explain why a 5% increase with respect to other studies was obtained here. Second, we should also point that Second, labelling a set of controlled sentences, read under optimal conditions, posed less of a challenge than labelling speech for telephone conversational speech (as in Yoon et al.'s study) or in read speech by professional announcers (as in Syrdal and McGory's study).

The Online Kappa Calculator (Randolph, 2008) was used to calculate the Fleiss kappa statistical measure (Yoon et al., 2004). This tool provides two variations of kappa: fixed-marginal multirater kappa and Randolph's free-marginal multirater kappa (Randolph, 2005; Warrens, 2010). Since raters knew previously about the presence of an intonational phrase and a nuclear accent for each sentence, fixed-marginal kappa was used. The fixed marginal kappa statistic obtained for the choice of phrasing levels was of 0.90 for Catalan, 0.90 for English, and 0.97 for Spanish, while the choice of prominence level had a kappa statistic of 0.87 for Catalan, 0.89 for English, and 0.93 for Spanish, indicating that those categories were reliably labeled. All in all, the results of the reliability test revealed that we can be confident about the reliability of the prosodic transcriptions in our data.

## 2.5. Rhythm metrics

In this section we will describe the rhythm metrics that were applied to the data. After data segmentation and prosodic labeling, we extracted vocalic and consonantal intervals and applied a selection of the most widely accepted rhythm metrics: %*V*,  $\Delta V$ , and  $\Delta C$  following Ramus et al. (1999), nPVI-V and rPVI-V following Grabe and Low (2002), and VarcoC and VarcoV following Dellwo (2006)

and Ferragne and Pellegrino (2004). In general, normalized metrics perform better than non normalized metrics for vocalic variability, whereas the reverse is the case for consonantal variability. And in general vocalic measures discriminate better than consonantal measures (see also White and Mattys, 2007a; Payne et al., in press, for a comprehensive review). Below we explain how these measures are calculated.

2.5.1. *Interval measures, which can be raw or normalized, are calculated as follows*

2.5.1.1. *Raw interval measures %V,  $\Delta V$ , and  $\Delta C$ .* Following the insight that infants perceive speech as a succession of vowels of alternating durations and intensities, alternating with periods of unanalyzed noise (i.e., consonants), Ramus et al. (1999) developed three measures of utterance rhythm that were based on the raw duration of vocalic and consonantal intervals in the sentence, as follows:

- %*V*: the proportion of vocalic intervals within the sentence (that is, the sum of the total duration of the vowels in the sentence divided by the total duration of the sentence).
- $\Delta V$ : the standard deviation of the duration of vocalic intervals within each sentence.
- $\Delta C$ : the standard deviation of the duration of consonantal intervals within each sentence.

Application of these metrics by Ramus et al. (1999) and Ramus et al. (2003) to languages of different rhythmic categories revealed a combination of  $\Delta C$  and %*V* to be the most useful in discriminating across categories.

2.5.1.2. *Rate-normalized measures: VarcoV and VarcoC.* Evidence that  $\Delta V$  and  $\Delta C$  were inversely proportional to speech rate (e.g., Dellwo and Wagner, 2003) as well as criticisms by Low et al. (2000) and Grabe and Low (2002) of the standard deviation measures proposed by Ramus et al. (1999) led these several authors to propose rate-normalized rhythm measures. Specifically, Dellwo (2006) proposed the normalized version of the consonantal standard deviation measure, namely, VarcoC. Later, Ferragne and Pellegrino (2004) and White and Mattys (2007a, 2007b) developed and tested VarcoV for their investigations. These measures are obtained as follows:

- VarcoV: standard deviation of vocalic interval duration divided by mean vocalic duration (and multiplied by 100).
- VarcoC: standard deviation of consonantal interval duration divided by mean consonantal duration (and multiplied by 100).

## 2.5.2. Pairwise variability metrics

A different approach to measuring rhythmic differences across languages was proposed by Low et al. (2000) and Grabe and Low (2002), who took the sequential



relationships between units in speech into account in quantifying variability in vocalic and consonantal intervals. The Pairwise Variability Index (PVI) measures express the average difference between adjacent units such as vowels, consonantal intervals, syllables, or feet (see also [Asu and Nolan, 2006](#)).<sup>9</sup> This approach was motivated by the observation that “stress-timed” languages exhibit more vocalic variability than “syllable-timed” languages.

The PVI-V, for instance, is calculated as the mean of the differences between successive intervals, and it can be normalized (nPVI) for speech rate variation by dividing it by the sum of intervals. The non-normalized and normalized versions are calculated as follows:

- rPVI-V, or non-normalized Pairwise Variability Index: the mean of the duration differences between successive intervals ( $V_s$ )
- nPVI-V, or vocalic normalized Pairwise Variability Index: the mean of the duration differences between successive intervals ( $V_s$ ) divided by the sum of the same intervals

## 2.6. Statistical analyses

All rhythmic and durational measures in this paper were analyzed using a mixed effects model (Generalized Linear Mixed Model, GLMM ANOVA) using IBM SPSS Statistics 19.0 (IBM Corporation, 2010). As [Baayen et al. \(2008\)](#) point out, generalized linear mixed models have the ability to account for random subject and item effects in one step of analysis and offer considerable advantages over repeated measures ANOVAs. First, to test the effects of syllable structure and language, a GLMM analysis was conducted separately on each of the rhythmic metrics (namely, the interval measures  $\%V$ ,  $\Delta V$ ,  $\Delta C$ , the Varco measures VarcoV and VarcoC, and the PVI measures for vowels nPVI and rPVI). The fixed independent variables were Syllable Type (three levels: open syllables, closed syllables, mixed) and Language (Catalan, English, Spanish). In all GLMM analyses, both subject and test item were set as crossed random effects. To test the effects of prominence level and phrasal position on syllable timing patterns, we carried out two GLMM analyses. In both cases, the response variable was syllable duration and the two fixed factors were Prominence Level (4 levels: unstressed, stressed unaccented, stressed accented, and nuclear accented) and Language (Catalan, English, Spanish) in one case, and Phrasal Position (3 levels: phrase-medial, intermediate-phrase-final, intonational-phrase-final) and Language (Catalan, English, Spanish) in the other. Again, both subject and test items were set as crossed random factors.

<sup>9</sup> [Low et al. \(2000\)](#) applied the nPVI to data from ten speakers of British English (stress-timed) and ten speakers of Singapore English (syllable-timed). The PVI results provided a robust acoustic basis for the impression of syllable timing in Singapore English since vowel durations were significantly more variable in British English than in Singapore English.

## 3. Results

The first part of this section presents the results of the potential effects of syllable structure complexity on the rhythm measures obtained across languages. The second part presents the results of the potential effects of levels of prominence and phrasal position on syllable durations across languages. Error bar graphs were computed which show the distribution of each of the dependent variables across the three conditions (CV-syllables, CVC-syllables, mixed) for all three languages.

### 3.1. The effects of syllable structure on rhythm metrics

As mentioned above, [Ramus et al. \(1999\)](#) argued that  $\%V$  was the measure that discriminates most successfully between perceived rhythmic classes. Since complexity in syllable structure has been claimed to be one of the most important structural factors underlying rhythmic class distinctions ([Ramus et al., 1999, p. 289](#)), one would expect this measure, together with the standard deviation for vocalic and consonantal intervals, to correlate closely with differences in syllable structure complexity. The question arises, then, as to whether differences in  $\%V$  are eliminated by controlling for syllable complexity in the speech materials.

The error bar graph in [Fig. 2](#) shows the mean  $\%V$  values for Catalan (black dots), English (dark grey dots), and Spanish (light grey dots). 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). As expected, the graph shows very clear effects of syllable structure on  $\%V$  measures. For all three languages, the simpler the syllable

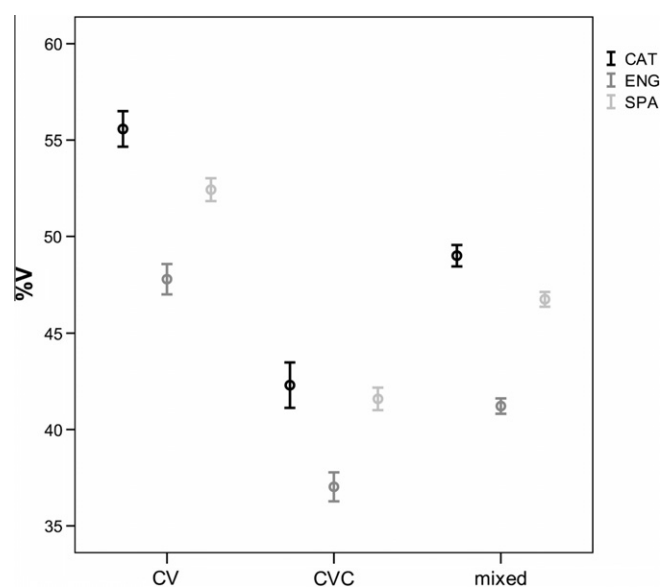


Fig. 2. Mean  $\%V$  values (in percentages) as a function of the phonotactic factor (CV, CVC, mixed sentences), for Catalan (black dots), English (dark grey dots), and Spanish (light grey dots). The bars represent the standard error of the mean.

structure, the higher the %*V*. That is, while the materials composed of predominantly open syllables range from 47.77% (English) to 55.57% (Catalan), the materials made up of predominantly closed syllables range from 37.02% (English) to 41.58% (Spanish), and the mixed materials lie in the middle range. However, and crucially, we find that the behavior of the three languages also differ in a systematic way within the three syllabic conditions. In two of the three conditions (CVC and mixed), the Catalan and Spanish data tend to cluster together, the two of them having a higher %*V* than English. This is a first indication that even though rhythmic distinctions as expressed by %*V* are partly dependent on the phonotactic properties of the speech materials, crosslinguistic differences can also be captured by this metric when syllable structure is controlled for.

A GLMM analysis was conducted with %*V* as the dependent variable, Syllable Type (3 levels: open, closed, mixed syllables) and Language (3 levels: Catalan, Spanish, and English) as fixed factors, and subject and item as crossed random factors. The analysis revealed a significant main effect of Syllable Type ( $F_{2,711} = 131.183$ ,  $p < .001$ ) and Language ( $F_{2,711} = 65.871$ ,  $p < .001$ ), and no significant interaction between the two factors ( $F_{4,711} = 1.021$ ,  $p = .396$ ) was found. Thus, our %*V* results show that, as expected, though the %*V* distinction is clearly dependent on syllable structure to some degree, it is also clear that even when syllable structure is controlled for, some crosslinguistic differences remain. The results show we should be cautious in taking the absolute %*V* values as strict points of comparison across languages, because they are at least partly dependent on the specific phonotactic properties of the materials being selected. Thus phonological control and the representativeness of the materials used are crucial considerations if we wish to compare languages using this index. We discuss this issue later in the Section 4.

Let us now analyze the behavior of the other two interval measures proposed by Ramus et al. (1999), namely,  $\Delta V$  and  $\Delta C$ , or standard deviation of *V* and *C*. The two error bar graphs in Fig. 3 show the mean of the standard deviation of the vocalic intervals,  $\Delta V$ , and the mean of the standard deviation of consonantal intervals,  $\Delta C$ , for all three languages (Catalan is represented by black dots, English by dark grey dots, and Spanish by light grey dots). Again, the data is separated according to predominantly CV-type utterances (left), predominantly CVC-type utterances (middle), and mixed utterances (right). A comparison between the two graphs reveals a very different behavior between  $\Delta V$  and  $\Delta C$ .  $\Delta V$  values, like %*V* values, clearly distinguish between Catalan and Spanish on the one hand and English on the other, in all three conditions, with English values being consistently higher. By contrast,  $\Delta C$  values only distinguish between these two groups of languages in the mixed condition.

A GLMM analysis was conducted with  $\Delta V$  as the dependent variable, Syllable Type (3 levels: open, closed, mixed syllables), and Language (3 levels: Catalan, Spanish, and English) as fixed factors, and subject and item as

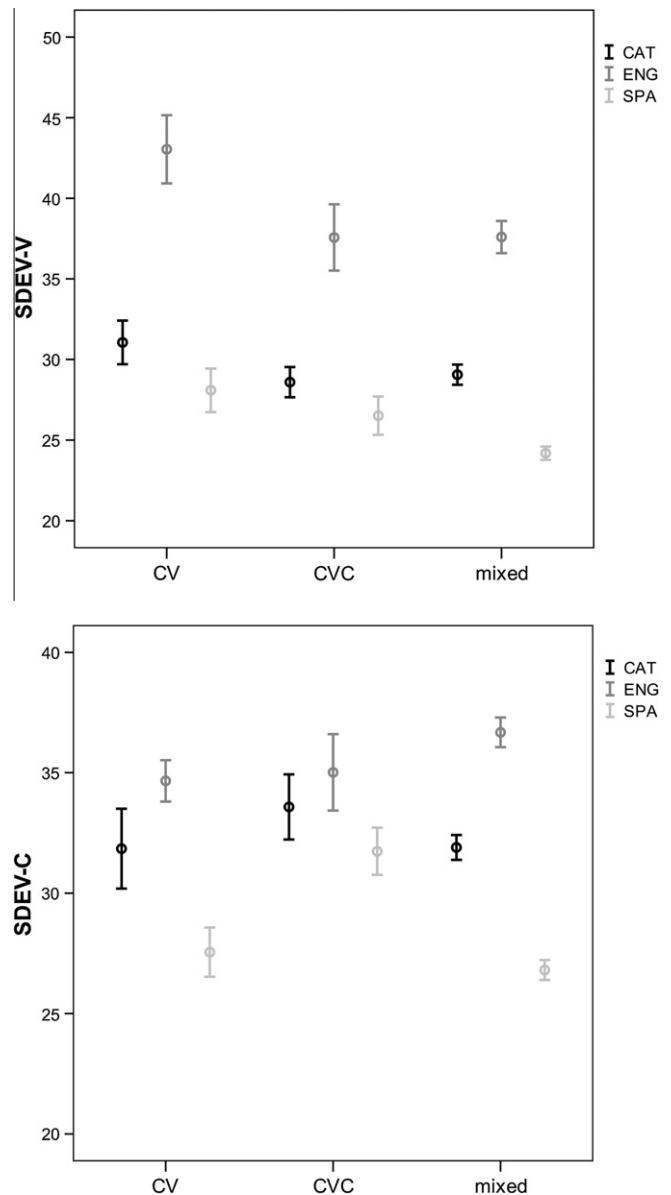


Fig. 3. Mean  $\Delta V$  (or standard deviation of *V*; upper graph) and  $\Delta C$  values (or standard deviation of *C*; bottom graph) as a function of the phonotactic factor (CV, CVC, mixed sentences), for Catalan (black dots), English (dark grey dots), and Spanish (light grey dots). The bars represent the standard error of the mean.

crossed random factors. The analysis revealed a significant main effect of Syllable Type ( $F_{2,711} = 7.818$ ,  $p < .001$ ) and Language ( $F_{2,711} = 84.902$ ,  $p < .001$ ), and no significant interaction between the two factors ( $F_{4,711} = 0.947$ ,  $p = .436$ ) was found. For  $\Delta C$  we found a main effect of Syllable Type ( $F_{2,711} = 3.163$ ,  $p < .05$ ) and Language ( $F_{2,711} = 36.058$ ,  $p < .001$ ), and also a significant interaction between Language and Syllable Type ( $F_{2,711} = 3.632$ ;  $p < .005$ ). Further pairwise comparisons on the  $\Delta C$  data revealed a significant difference between all pairs of languages only in the mixed condition (at  $p < .001$ ), revealing that  $\Delta C$  only discriminates consistently between languages in the mixed condition and is thus strongly dependent on the phonotactic properties of the speech materials.

The results for the % $V$  and standard deviation data constitute a preliminary indication that differences in vocalic interval durations across languages are not only the reflection of the phonotactic differences across the languages involved and that the systematic differences across % $V$  and  $\Delta V$  values across different types of materials may be reflecting a systematic crosslinguistic difference in timing organization.

As noted above, objections on the part of certain authors to any reliance on raw standard deviation measures  $\Delta C$  and  $\Delta V$  on the grounds that they are speech rate dependent (Low et al., 2000; Grabe and Low, 2002; Dellwo and Wagner, 2003) led these authors to propose rhythmic measures that sought to correct for this dependency. The two error bar graphs in Fig. 3 show the mean VarcoV and VarcoC values, for all three languages (Catalan is represented by black dots, English by dark grey dots, and Spanish by light grey dots). 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). Importantly, VarcoV (but not VarcoC) is higher in English than in Catalan and Spanish, across all syllable types, even though in some cases the effects are not significant. The two graphs in Fig. 4 show a contrasting behavior between VarcoV and VarcoC that clearly resembles the contrasts reported between  $\Delta V$  and  $\Delta C$  in the preceding section. The VarcoV graphs show that the English data consistently display a significantly higher vowel variability than Catalan and Spanish, across the three conditions. By contrast, the VarcoC data do not capture any differences across languages, not even in the control mixed condition.

GLMM analyses were applied to the VarcoV and VarcoC data, with VarcoV and VarcoC as the dependent variable, Syllable Type (3 levels: open, closed, mixed syllables) and Language (3 levels: Catalan, Spanish, and English) as fixed factors, and subject and item as crossed random factors. The VarcoV analysis revealed significant effects of Language ( $F_{2,711} = 49.699$ ,  $p < .001$ ), no significant effects of Syllable Type ( $F_{2,711} = 1.312$ ,  $p = .270$ ), and no significant interaction between the two factors ( $F_{4,711} = 1.415$ ,  $p = .227$ ). By contrast, the VarcoC results revealed no effects of Syllable Type ( $F_{2,711} = 0.723$ ,  $p = .486$ ) nor of Language ( $F_{2,711} = 1.968$ ,  $p = .141$ ), and a significant interaction between the two factors ( $F_{2,711} = 5.496$ ,  $p < 0.001$ ). Further pair comparisons showed no significant differences between the three levels in the Syllable Type condition and a statistically significant difference between the English–Spanish pair (at  $p < 0.05$ ). Clearly, while VarcoV successfully discriminates between English vs. Catalan and Spanish in all the materials, this is not the case with VarcoC. Thus, VarcoV, in contrast with VarcoC, has a high discriminating power which is not dependent on syllable structure, and thus it can be considered a robust and reliable rhythm measure across languages. Moreover, VarcoV and VarcoC mirror the results reported for the raw vowel and consonantal standard deviation data,  $\Delta V$  and  $\Delta C$ .

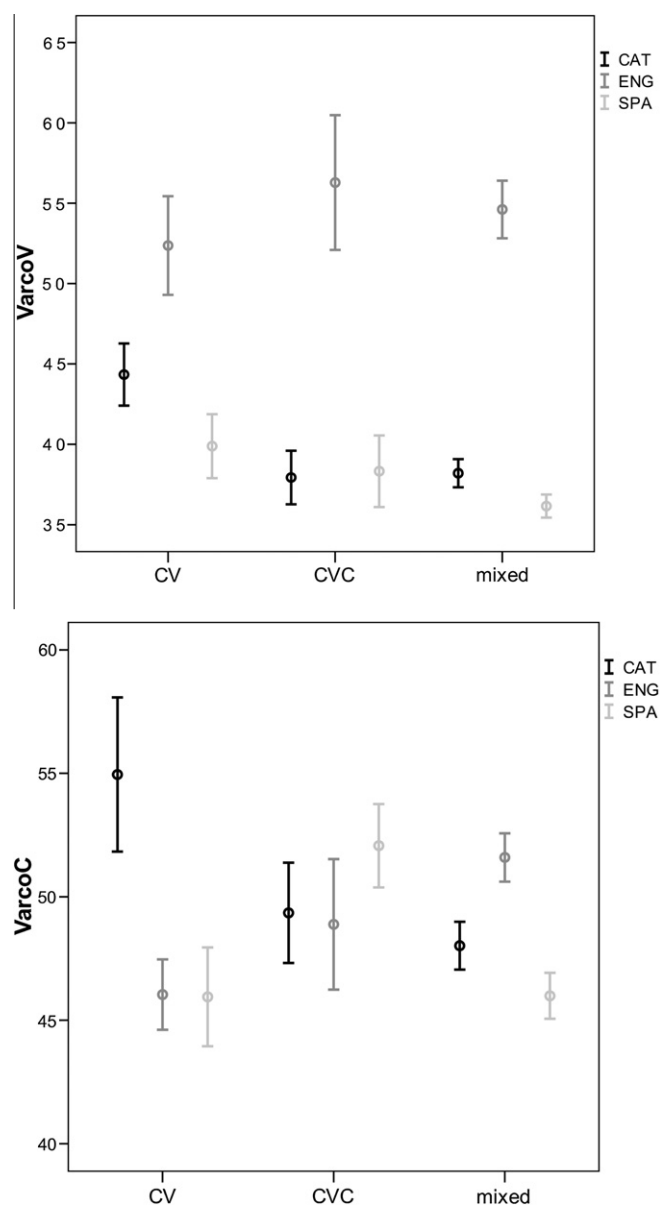


Fig. 4. Mean VarcoV (upper graph) and VarcoC values (bottom graph) as a function of the phonotactic factor (CV, CVC, mixed sentences), for all three languages. The bars represent the standard error of the mean.

With respect to the PVI measures, we expected to find a higher vocalic PVI value in English, reflecting more durational variability between successive vowels. Although Low et al. (2000) also proposed a consonantal PVI (PVI-C), they claimed that it was the vocalic PVI, and not the consonantal PVI, which was able to reflect the rhythmic difference between British English and Singapore English (see Low et al., 2000, p. 396–397). The two error bar graphs in Fig. 5 show the mean raw vocalic measures (rPVI-V, upper graph), and the normalized measures (nPVI-V, lower graph), for all three languages (Catalan is represented by black dots, English by dark grey dots, and Spanish by light grey dots). In the graphs, the data are separated into the three types of materials used, namely, predominantly CV-type utterances (left), predominantly CVC-type utterances

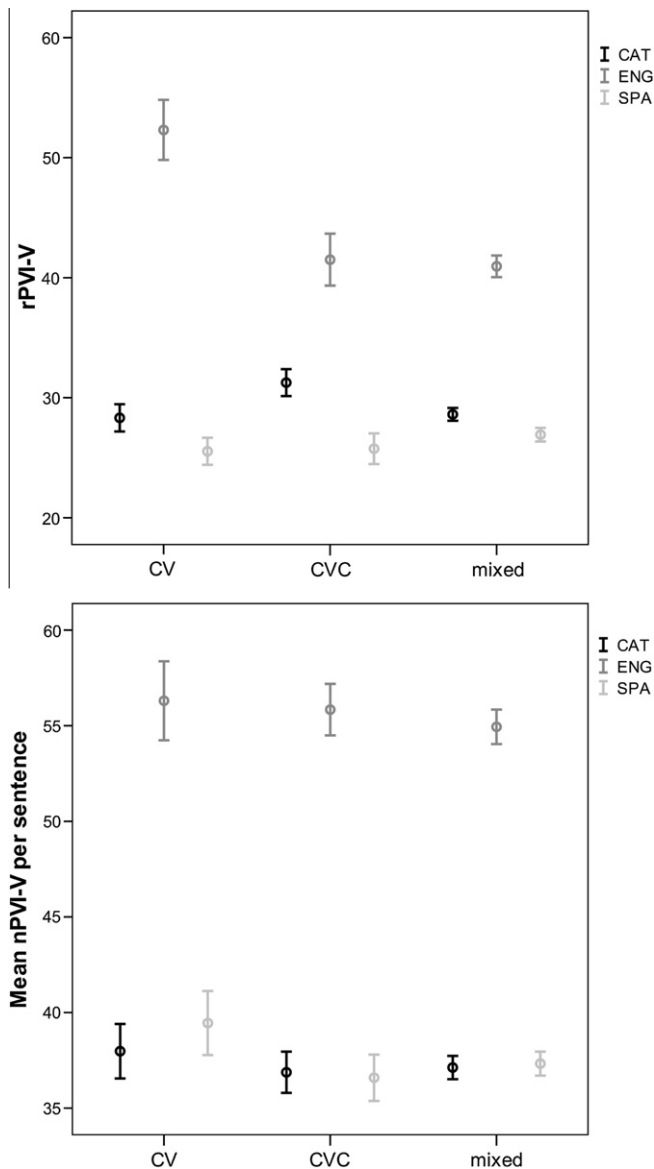


Fig. 5. Mean raw vocalic Pairwise Variability Index values (upper graph) and normalized PVI-V values (bottom graph) as a function of the phonotactic factor (CV, CVC, mixed sentences), for all three languages. The bars represent the standard error of the mean.

(middle), and mixed utterances (right), for Catalan (white boxes), English (striped boxes), and Spanish (grey boxes). As expected, English speakers displayed consistently higher PVI-V values (both raw and normalized) compared to Spanish and Catalan speakers, across all three conditions. Thus English exhibits consistently greater variability between successive vowel durations than Spanish and Catalan. As for the effects of syllable structure, neither the raw nor the normalized PVI measures show much effect of this factor. The PVI-V data thus mirror the behavior of  $\Delta V$  and VarcoV measures in the preceding sections, as they are relatively independent of syllable structure and do reflect stable durational differences across languages.

As expected, the results of a GLMM analysis of variance revealed a distinct behavior between the rPVI-V and

the nPVI-V measures. On the one hand, the raw PVI data revealed a significant main effect of both Syllable Type ( $F_{2,711} = 5.775$ ,  $p < .005$ ) and Language ( $F_{2,711} = 187.005$ ,  $p < .001$ ), and a significant interaction between Language and Syllable Type ( $F_{2,711} = 10.277$ ,  $p < .001$ ). On the other hand, the normalized PVI analysis showed a significant main effect of Language ( $F_{2,711} = 203.643$ ,  $p < .001$ ), no effects of Syllable Type ( $F_{2,711} = 1.224$ ,  $p = .295$ ), and no significant interaction between Language and Syllable Type ( $F_{4,711} = 0.232$ ,  $p = .921$ ), indicating that this measure is more robust in capturing rhythmic distinctions across phonotactic conditions.

To summarize, our results in this section show that various vocalic measurements, namely, %V,  $\Delta V$ , VarcoV, and nPVI-V, provide us with a more robust index of rhythmic classes than consonantal measures. Measures of consonantal interval variability such as  $\Delta C$  and VarcoC are more sensitive to phonotactic differences than to rhythmic differences. This constitutes a preliminary indication that vocalic measures constitute a more stable set of measures across languages. However, while %V is indeed confounded by syllable structure differences in the materials, the other three measures are more robust and relatively independent of syllable structure. This contrast between the vocalic rhythmic metrics may be due the fact that  $\Delta V$ , VarcoV, and nPVI-V are all measures of *variability* in vocalic interval duration, while %V is an expression of overall vocalic duration. Also, the cross-linguistic differences are most apparent when variability is quantified through pairwise comparisons (PVI-V).

An important finding from this first study is that the majority of rhythmic indices reflect a systematic difference between English on the one hand and Catalan/Spanish on the other, regardless of the syllable structure types present in the language materials. What is the main acoustic correlate for this consistent distinction between rhythm indices across the two groups of languages? We have seen that phonotactic differences cannot exclusively be at the root of such distinction. What about vowel reduction properties? In our data, even though both Catalan and English have a phonological process of vowel reduction they are unexpectedly classed as two languages with different rhythmic properties. In the next section we will test whether rhythmic differences in our three languages are correlated with two timing phenomena operating at the phrasal level, the highest level in the prosodic hierarchy, namely, the implementation of durational phrasal prominence in prosodic heads and the implementation of durational lengthening at edges of prosodic domains.

### 3.2. Effects of prominence level and phrasal position on timing patterns

In this section we describe the effects of phrasal prominence and phrasal position on the timing patterns found in our data.



### 3.2.1. Effects of prominence levels on syllable duration

In our data, phrasal prominence was labeled using the following four levels: unstressed, lexically stressed but unaccented, accented, and nuclear accented. We hypothesized that different levels of prominence would be reflected in durational differences in our data, with stronger effects for English than for Catalan and Spanish (Ortega-Llebaria and Prieto, 2007 for Spanish, Astruc and Prieto, 2006 for Catalan, and Turk and White, 1999 for English).

The error bar graph in Fig. 6 shows the mean syllable duration (in ms.) for the three languages (Catalan = black, English = dark grey, Spanish = light grey), as a function of prominence. In the x axis, the data are separated into unstressed unaccented positions, stressed unaccented positions, stressed accented positions, and nuclear accented positions. Results reveal regular and consistent patterns of behavior across languages and across conditions. As expected, the extra amount of lengthening associated with accented and nuclear accented syllables is much larger in English than in Spanish or Catalan. With respect to the differences between Catalan and Spanish, we observe a tendency for Catalan to show more lengthening in both accented positions and the nuclear stress positions. Interestingly, the fact that the unstressed syllables in the three languages are more or less equal in duration indicates that there is in fact no clear evidence for compression of unstressed syllables in English.

A GLMM analysis was applied to the data. The response variable was syllable duration (in ms.), and the main factors were prominence level (4 levels: unstressed, stressed unaccented, stressed accented, and nuclear accented) and Language (3 levels: Catalan, English, and

Spanish). Results revealed a significant main effect of Language ( $F(1,2) = 114.940$ ;  $p < .001$ ) and prominence level ( $F(1,3) = 864.955$ ;  $p < .001$ ) on syllable duration and a significant interaction between Language and prominence level ( $F(1,6) = 52.534$ ;  $p < 0.001$ ). Pairwise comparisons revealed that the three languages differ in their durational patterns (at  $p < .001$ ) and that different prominence levels are significantly different (at  $p < .001$ ), with the exception of the distinction between unstressed unaccented and stressed unaccented.

The results reported in this section reveal a language-specific pattern in the phonetic realization of syllable duration across different levels of prominence. We found that English has a stronger durational marking of different levels of prominence. This confirms previous qualitative observations on the difference between English and Spanish (Hualde, 2005, p. 273)<sup>10</sup>. With respect to the difference between Spanish and Catalan, we found that Catalan has a slight tendency towards stronger marking of prosodic heads than Spanish. This difference can probably be related to the presence of vowel reduction, as claimed by Ortega-Llebaria and Prieto (2010). They reported that stressed and unstressed syllables in both Catalan and Spanish have similar durations across conditions, regardless of the fact that Catalan has a pervasive vowel reduction process in unstressed position. Yet they also reported larger duration differences between stressed and unstressed syllables in the [a]-[ə] alternation. They argue that the changes in vowel quality between corresponding vowels not only provide a spectral correlate to the stress contrast but also have the effect of amplifying any duration differences that are triggered by stress. Interestingly, the same pattern of results was found by Fant et al. (1991a, p. 360) in their study of the correlates of stress in Swedish, English, and French. They concluded that “English and Swedish are comparable in the sense that the stressed syllable attains a stress-induced lengthening on the order of 100–150 ms, in positions other than before a pause, while in French the syllable lengthening is of the order of 50 ms only”.

The results in this section thus suggest that the different implementation of prominence across languages might be an important factor in the perception of rhythm distinctions. Clearly, the analysis of specific data on the durational implementation of prominence across languages might be a useful tool that can complement rhythm indices in cases of mixed languages like Catalan, as it allows us to investigate the acoustic basis of the rhythmic distinctions in a fine-grained way.

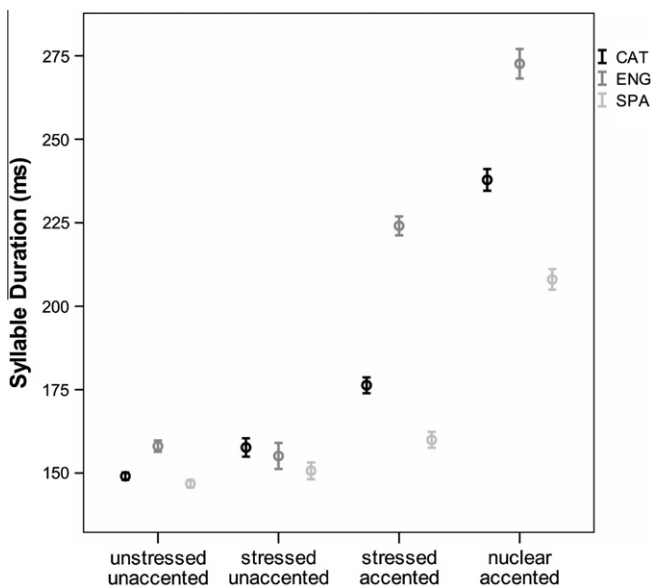


Fig. 6. Mean syllable duration (in ms) in the three languages (Catalan = black, English = dark grey, Spanish = light grey). The data are separated into unstressed positions, unstressed accented positions, stressed accented positions, and nuclear accented positions. The bars represent the standard error of the mean.

<sup>10</sup> As Hualde (2005, p. 273) points out, “The differences in duration between stressed and unstressed syllables are, however, much greater in English than in Spanish. As we saw in Section 1.3.2, segmental effects on the quality of both consonants and vowels of both stressed and unstressed syllables are also much greater in English than in Spanish. In English the difference between stressed and unstressed syllables is thus more salient than in Spanish. [...] This contributes to a perceptual difference between the two languages, interpretable in terms of rhythm”.

### 3.2.2. Effects of phrasal position on syllable duration

Numerous studies have indicated that prosodic phrase boundaries are marked by a variety of acoustic phenomena including segmental lengthening (Wightman et al., 1992 and Yoon et al., 2007 for English, among others). It has not been firmly established, however, whether this phenomenon has different patterns of implementation across languages and whether it can be related to the establishment of rhythmic classes.

Fig. 7 shows the mean syllable duration (in ms.) for the three languages (Catalan = black, English = dark grey, Spanish = light grey), as a function of phrasal position. The data are now separated into the following phrasal positions: non-final (left), end of intermediate phrase (middle), and end of intonational phrase (right) – see Section 1.4 for a description. First, the data reveal that the three languages display 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 either an *ip* or an *IP*. On the other hand, though the differences between Catalan and Spanish are smaller, Catalan has longer syllables than Spanish in those positions. All in all, the three languages show strong differences between duration patterns at the edges of an *ip* relative to those occurring at the edges of an *IP*.

We carried out a GLMM analysis with this data. The response variable was Syllable Duration and the two fixed factors were Phrasal Position (3 levels: phrase-medial, intermediate-phrase final, intonational-phrase final) and

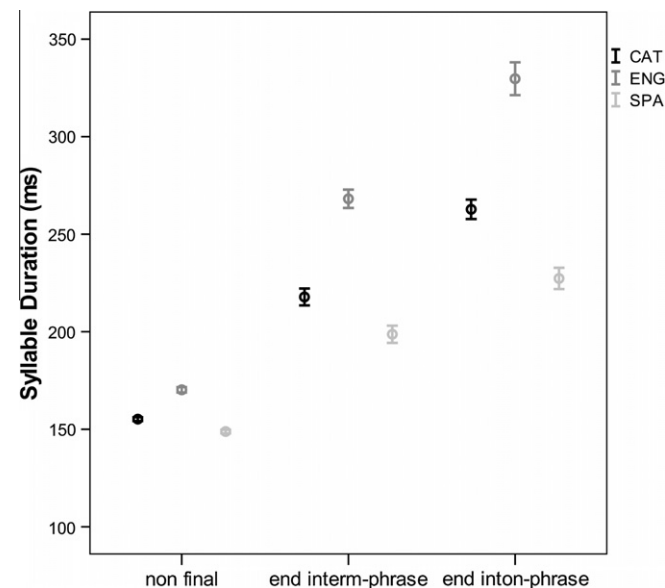


Fig. 7. Mean syllable duration (in ms.) in the three languages (Catalan = black, English = dark grey, Spanish = light grey). The data are separated into the following phrasal positions: non-final (left), end of intermediate phrase (middle), and end of intonational phrase (right). The bars represent the standard error of the mean.

Language (Catalan, English, Spanish). Again, both subject and test items were set as crossed random factors. Results revealed a significant main effect of Language ( $F(1,2) = 230.442$ ;  $p < .001$ ) and Phrasal Position on syllable duration ( $F(1,2) = 1409.839$ ;  $p < .001$ ) and a significant interaction between Language and Phrasal Position ( $F(1,4) = 59.897$ ;  $p < .001$ ). Pairwise comparisons revealed that the three languages differ in their durational patterns (at  $p < .001$ ) and that the three phrasal positions are significantly different (at  $p < .001$ ).

The data in this section appears to reveal language-specific differences in the implementation patterns of final lengthening across the three languages. English appears to have a stronger pre-boundary lengthening effect than Spanish and Catalan. Catalan also seems to have stronger marking of prosodic boundaries. These findings also provide evidence that the established distinction in phrasing (namely, the *ip* and *IP*) has clear durational correlates in the three languages. This confirms recent results by Yoon et al. (2007) for American English and likewise would seem to justify the level distinction made by the ToBI transcription systems for Catalan and Spanish (e.g., Cat\_ToBI: Prieto, in press; and Sp\_ToBI: Beckman et al., 2002; Estebas-Vilaplana and Prieto, 2010). In the remainder of this paper, we will discuss the results of the present investigation in the light of the established rhythmic distinctions across languages.

### 3.2.3. A general linear model

In order to appreciate the joint effects of all factors, an GLMM analysis was applied to the syllable duration data, now taking into account all the fixed factors that were controlled for in the production experiment, namely, syllable type (3 levels: open, closed, and mixed), prominence level (4 levels: unstressed, stressed unaccented, stressed accented, and nuclear accented), phrasal position (3 levels: phrase-medial, intermediate-phrase final, intonational-phrase final), and language (3 levels: Catalan, English, and Spanish). Again, subjects and item were set as crossed random factors. As expected, the four fixed factors were found to have a significant main effect on syllable duration, namely syllable structure ( $F(1,2) = 98.247$ ;  $p < .001$ ), prominence level ( $F(1,3) = 212.439$ ;  $p < .001$ ), phrasal position, ( $F(1,2) = 526.258$ ;  $p < .001$ ) Language ( $F(1,2) = 80.100$ ;  $p < .001$ ). Given the absolute  $F$  values, the size effects are shown to be larger for phrasal position and prominence level, with smaller size effects for language and syllable structure.

Pairwise analyses showed a significant interaction between Language and Syllable Type ( $F(1,4) = 6.058$ ,  $p < .001$ ), Language and Prominence Level ( $F(1,6) = 23.829$ ,  $p < .001$ ), and Language and Phrasal Position ( $F(1,4) = 23.679$ ;  $p < .001$ ), as it was expected and reported in the separate abovementioned GLMM analyses. These two-way interactions underscore the importance of specific differences in timing phenomena across the three languages. No significant interaction was

obtained between Syllable Type and Phrasal Position ( $F(1,4) = 2.653$ ,  $p = .031$ ), meaning that there is a consistent effect of Phrasal position across all levels of Syllable Structure. By contrast, a significant interaction was found between Syllable Type and Prominence Level ( $F(1,6) = 7.027$ ,  $p < .001$ ) and between Prominence Level and Phrasal Position ( $F(1,2) = 8.309$ ,  $p < .001$ ). The first interaction can be explained by the fact that in the case of unstressed and stressed unaccented syllables, the effects of syllable type on duration are not as strong as in other prominence positions. It can be argued that low-level prominence positions tend to mask the duration differences between CV and CVC syllable types. In the second case, there were no cases of stressed unaccented and stressed accented syllables at the ends of ips and/or IPs, since all syllables in such positions were classified as either unstressed or stressed nuclear accent.

With respect to three-way interactions, they were all significant (at  $p < .05$ ) with the exception of Language\*Prominence Level\*Phrasal Position. First, Language\*Syllable Type\*Prominence Level ( $F(1,12) = 2.358$ ;  $p < .05$ ) and Syllable Type\*Prominence Level\*Phrasal Position ( $F(1,4) = 8.770$ ;  $p < .001$ ) were significant three-way interactions. As it was mentioned before, unstressed and stressed unaccented syllables do not show the same patterns of duration differences across syllable types and across languages, since these two types of syllables tend to be much more similar in duration values than at other prominence levels, and this explains the three-way interactions with the factor Prominence Level. With respect to the triple interaction between Language\*Syllable Type\*Phrasal Position ( $F(1,8) = 3.382$ ;  $p < .05$ ), syllables in phrase-medial positions tend to be more neutralized in duration and do not show the same patterns that can be found in syllables in phrase-final positions (be it at the end of intermediate phrases or the end of intonational phrases). Finally, the significant four-way interaction between the four fixed factors ( $F(1,8) = 3.290$ ;  $p < .05$ ) can be explained by the weak magnitude effect of sentence-medial and unstressed syllables on duration when compared to the other position and prominence categories.

## 4. Discussion

### 4.1. Effects of syllable structure on rhythm measures

The results from the production experiment presented in this paper have revealed significant effects of syllable structure differences on several rhythm measures, namely, % $V$ ,  $\Delta V$ ,  $\Delta C$ , VarcoC, and rPVI. This result came as no surprise since many authors had previously pointed out that interval metrics were strongly influenced by syllable structure (see Ramus et al., 1999, among others), and indeed the metrics were in part developed precisely to reflect differences in this structure. As Ramus et al. (1999, p. 273–274) points out, “ $\Delta C$  and % $V$  appear to be directly related to syllabic structure”. The fact that these rhythmic indices are highly

dependent on the phonotactics of the materials being used means that one should be cautious in the methodological procedure used to select those materials. The representativeness of the materials should be controlled so that it matches with counts of the phonotactic distributions in the target language, ideally through independent measures that quantify typical syllable complexity across languages. We believe that the question of the representativeness of the target materials is an important one in rhythm studies.

Importantly, the normalized measures of vocalic variability, VarcoV and nPVI-V – that is, the normalized measures of vocalic variability<sup>11</sup> – showed no effects of syllable structure and a clear effect of language. The fact that these indexical measures are independent from phonotactic factors but nevertheless reveal crosslinguistic differences is a clear indication that additional factors other than syllable structure are playing an important role in the rhythmic distinctions being investigated. This is reinforced by the fact that the other vocalic metrics investigated, namely, % $V$ ,  $\Delta V$ , and rPVI, despite showing effects of syllable structure, nevertheless reveal differences between English and Catalan/Spanish when syllable structure is controlled for. Thus the picture that emerges from the data has two main aspects. First, durational variability differences exist across languages, over and above those caused by differences in syllable structure. Second, we find three types of metrics depending on their sensitivity to phonotactic and/or prosodic factors, as follows: (i) metrics that are purely dependent on syllable structure such as  $\Delta C$  and VarcoC; (ii) metrics that are purely dependent on phrasal prosodic factors (VarcoV and nPVI-V); and (iii) metrics that are dependent on both phonotactic and prosodic factors (% $V$ ,  $\Delta V$ , and rPVI).

Also importantly, there is a consistent difference between the behavior of vocalic and consonantal measures of rhythm. Crucially, the consonant interval metrics investigated, whether normalized or non-normalized ( $\Delta C$  and VarcoC), were not able to discriminate across languages (except for  $\Delta C$  in the mixed condition). The low discriminatory power of the consonantal indices indicates that these measures are highly dependent on phonotactic factors, which constitute just one of several elements contributing to the percept of rhythm. This finding is compatible with previous reports on consonantal metrics. Not surprisingly, White and Mattys (2007a, b) found that the vocalic measures, namely, VarcoV, nPVI-V, and also % $V$ , offered the most effective analyses in capturing rhythmic differences in first and second language learners. Yet measures of consonantal interval variation ( $\Delta C$  and VarcoC) proved to be far less effective or consistent in discriminating between rhythmic types. By contrast, the high discriminatory power of vocalic measures (both interval measures

<sup>11</sup> Remember that VarcoV and nPVI-V are the non-normalized counterparts to  $\Delta V$  and the raw PVI-V respectively.

and standard deviation measures) would make sense because the vowel portion of the string is known to better reflect the timing patterns in language (Low et al., 2000; Grabe and Low, 2002, and others) and therefore be more sensitive to differences in these patterns. Moreover, the high informational load encoded by vowels is an essential part of Ramus et al. (1999) that children's early perception of speech is primarily centered on vowels.

The results reported in this article appear to have implications for the view that phonotactic structure forms the basis of the rhythmic distinctions across languages (Ramus et al., 1999). Indeed, this view of rhythm would predict that rhythmic differences should be minimized when linguistic materials are controlled for syllable structure. Yet consistent crosslinguistic differences were obtained in our experiment. Particularly telling in this respect are the results from cross-dialectal studies that have begun to uncover differences in rhythm within a given language (Low et al. (2000) for the Singapore and British varieties of English, Frota and Vigário (2001) for the European and Brazilian varieties of Portuguese, O'Rourke (2008) for the Lima and Cuzco varieties of Spanish, Nolan and Asu (2009) for Peninsular Spanish and Mexican Spanish, and White et al., 2009 for northern and southern varieties of Italian, among others). Though many of these dialectal varieties share key phonotactic properties like syllable structure and vowel reduction, they display important rhythmic differences, as evident from both the metrics and perceptual evaluation. For example, Low et al. (2000) report a robust acoustic difference between the British ("stress-timed") and Singapore ("syllable-timed") varieties of English, regardless of the fact that syllable structure and vowel reduction properties are similar across the two varieties. Similarly, O'Rourke (2008) shows that, even though Spanish has been taken as a prototypical "syllable-timed" language, a wider range of possible speech rhythms have been observed for two varieties of Peruvian Spanish, namely, Lima and Cuzco Spanish, pointing to the need for further cross-dialectal research. Perhaps even more striking is the finding that rhythmic indices can vary systematically between two different speech styles of the same language variety (see Payne et al., 2009, 2010, for a comparison of the rhythm of child-directed and adult-directed speech for English, Catalan, and Spanish).

The case of Catalan represents another argument against the view that vowel reduction and phonotactic properties are the key properties of the rhythm percept. As stated by several authors, the mixed phonological properties of this language would predict that it should behave as an intermediate language with respect to rhythm (see Nespó, 1990, among others). As mentioned above, syllable structure in Catalan is more complex than in Spanish and it displays a generalized process of vowel reduction which affects all vowels except for [i] and [u]. Yet independent experimental evidence (and our own results) clearly clusters Catalan with "syllable-timed" languages like Spanish. This is a clear indication that vowel reduction in this

language is not correlated with durational reduction, as has been argued in previous research (Ortega-Llebaria and Prieto, 2010). As we will see below, the case of Catalan suggests that durational variability cannot always be traced back to phonological properties of the language.

#### 4.2. *The status of Catalan*

Although Catalan has been classified as an intermediate language with respect to rhythm, linguists have not yet reached a firm consensus on this status. In our data, rhythm metrics do not behave consistently with respect to the classification of Catalan. Though some rhythmic scores revealed differences between the English data on the one hand and the Catalan/Spanish data on the other (namely, nPVI-V and VarcoV), other measures also indicated smaller yet still significant differences between Catalan and Spanish (namely, %V and  $\Delta V$ ). Similarly, the analysis of adult-directed speech carried out by Payne et al. (in press) showed that Catalan is distinct from English in all of the scores but is distinct from Spanish only in the normalized vocalic metrics (VarcoV and nPVI-V), and similar for all other metrics. Gavaldà-Ferré's (2007) comparison of Catalan with other languages included in the BonnTempo corpus (Dellwo et al., 2004) found that data for %V and  $\Delta C$  classify Catalan as a "syllable-timed" language whereas results for vocalic nPVI-V and rPVI-V characterize Catalan as an intermediate language. Yet even though the rhythm indices that are currently available are not in complete agreement where Catalan is concerned, our quantitative results taken together tend to agree with Ramus et al. (1999) that Catalan tends to group quantitatively with the other "syllable-timed" languages like French and Italian.

There is thus little evidence from these metrics to support the view that Catalan should belong to an intermediate rhythmic class. Though Catalan displays a systematic process of vowel reduction and a higher syllabic complexity than Spanish, surprisingly no rhythm metrics consistently classify Catalan as a "stress-timed" language. Both Catalan and Spanish tend to cluster with syllable-time measures, even though some scores for Catalan are indeed significantly different from those for Spanish. Ramus et al. (2003) already stated that vowel reduction in Catalan is not enough to make it separate from "syllable-timed" languages. They reached the following conclusion: " $\Delta V$  does not separate Catalan from syllable-timed languages, suggesting that vowel reduction in Catalan does not quantitatively impact on this variable. This is consistent with our perceptual results which suggest that vowel reduction in Catalan is not enough to make it depart from syllable-timing" (Ramus et al., 2003, p. 5). In accordance with this, we claim that the "syllable-timed" behavior of Catalan is rooted in the way Catalan implements vowel reduction, which is very different from how it is implemented in English. In a recent study about the correlates of stress distinctions in Catalan and Spanish, Ortega-Llebaria and Prieto



(2010) show that although Catalan has vowel reduction in unstressed positions, this difference is reflected in a weak way in the duration patterns, which behave almost in the same way in the two languages. In essence, Catalan vowel reduction does not correlate very strongly with durational reduction in unstressed position (as it has similar timing properties to Spanish), but rather correlates primarily with vowel quality changes. A similar vowel reduction pattern was found in Singapore English. Low et al. (2000) compared spectral and durational patterns of reduced vowels in Singapore English and British English. The spectral results revealed that Singapore English has vowel reduction, but reduced vowels are less centralized in the F1/F2 space than reduced vowels in British English. Importantly, reduced vowels in Singapore English are also longer than their counterparts in British English.

Therefore, we claim that vowel reduction processes can be relatively independent of durational reduction. Under this view, it makes sense that a language like Spanish, with no phonological vowel reduction, has comparable duration differences between stressed and unstressed vowels to Catalan. Further evidence of the relative independence of vowel reduction and durational reduction is the experiment undertaken by Gavalda-Ferré (2007). Crucially, her data reveals a lack of rhythmic distinction between the Eastern Catalan dialect, with a full vowel reduction system, and the Western Catalan dialect, with a very restricted vowel reduction system. Thus in our view the languages of the world can implement two different types of vowel reduction: one that compresses syllable duration very strongly, such as that found in German or British/American English, and another that is purely qualitative and has little impact on vowel or syllable compression, like the cases of Catalan and Singapore English. Thus even though vowel reduction and durational reduction influence each other, they can also be controlled separately in speech production. This goes against the idea that vowel reduction processes should automatically lead to a percept of “stress-timed” rhythm.

Finally, let us return to the issue of the classification of the Catalan language in the hypothesized rhythmic scale. As noted above, we find partially contradictory evidence across the rhythm metrics in the classification of Catalan. Four production studies, apart from our own, have shown that Catalan behaves differently depending on what rhythm measures are taken (see Ramus et al., 1999; Low et al., 2000; Gavalda-Ferré, 2007; Payne et al., in press). Though the variable performance in the various metrics in the classification of Catalan might be due to the fact that different metrics are sensitive to different factors, these discrepancies suggest the need for an in-depth analysis of the timing patterns in this language. On the other hand, perhaps it is a premature enterprise to try to impose the “intermediate” category with respect to rhythm. After all, the grouping of languages into rhythmic classes is based on an auditory impression of languages as being more or less regularly timed and there are almost no perceptual studies correlat-

ing acoustic cues with rhythmic impressions.<sup>12</sup> What would the auditory impression of an intermediate language be? How can this be quantified perceptually? Since this question is not solved, we believe that it is probably too early to start looking for the acoustic correlates of such an intermediate rhythm.

A related issue that is worth mentioning is the existence of a categorical or gradient rhythmic distinction across languages, which has been a recurrent discussion in the rhythm literature. As mentioned above, rhythmically mixed languages have been considered to constitute key evidence in the debate between the existence of either categorically distinct rhythmic classes or a rhythmic continuum. While the research reported by Ramus et al. (1999), Low et al. (2000), and Grabe and Low (2002), using different quantitative measures, has shown that it is possible to achieve useful scalar characterization of the rhythm of different languages, it is still an open question whether we need to acknowledge the existence of a categorical rhythmic distinction between languages. If we use acoustic correlates of timing and durational variability, we might indeed find that differences across languages are of a continuous nature.<sup>13</sup> Yet the rhythm *percept* of these production differences might well be categorical in nature. Thus we need to investigate in greater depth the mapping between perceptual properties of rhythm on the one hand and the acoustic properties present in speech on the other (see Barry et al., 2009 for a recent study).

#### 4.3. *Durational marking of prosodic heads and edges*

The analysis of the durational patterns of accentual lengthening and final lengthening in the three languages under investigation appears to reveal that: (a) English appears to have a stronger durational marking of accentual prominence than Catalan and Spanish, and (b) English also appears to have a stronger durational marking of phrase-final lengthening than Catalan and Spanish. In this sense, our data have revealed clear crosslinguistic differences in the implementation of durational patterns marking both phrasal prominence and edge positions. Whether these two phenomena go hand in hand in languages remains an open question. However, for the languages investigated, since this grouping correlates with perceptual groupings for rhythm, the results suggest that these two processes might have a strong impact on the perception of rhythm.

The results reported in this article support the view that language-specific implementation of prosodic timing phenomena might be an integral part of the rhythmic differences across languages. As mentioned noted, this view is

<sup>12</sup> We would like to thank an anonymous reviewer for pointing this out to us.

<sup>13</sup> The considerable amount of work dealing with cross-dialectal differences in rhythm (see Carter 2005 or O'Rourke 2008, to name but two) has unveiled in many cases a gradient distinction between dialects of the same language, which might in some cases reveal a change in progress.

not new (see in particular Fant et al. (1991a,b) and Beckman (1992). In their study of the phonetics of stress in Swedish, English, and French, Fant et al. (1991a) claimed that the durational correlates of prominence could be a reflection of the rhythmic classification.<sup>14</sup> Beckman (1992) proposed the comparison of timing patterns across languages in order to find language-specific timing patterns of prosodic structure. We think that these initial results warrant the investigation of phrasal prosodic durational correlates across languages and the potential relationship of such correlates with perceived rhythm classes. One of the advantages of using phrasal timing measures is that they offer precise and direct tools that can be used to evaluate the discrepancies found between the rhythm metrics in the classification of languages. For example, the Catalan phrasal timing results demonstrate that the partially contradictory behavior of the rhythmic indices can be traced back to the tendency of Catalan to have slightly stronger durational marking of prosodic prominence and boundaries than Spanish.

The view of speech rhythm as (at least in part) a phrasal timing mechanism that is organized through prosodic structure is in direct accordance with recent experimental results on the performance of speakers of different languages in rhythmic entrainment. Recent investigations by Cummins and Port (1998) and Cummins (2002) have shown that while English speakers were able to produce a sequence of stressed syllables in time with a metronome, keeping time with the metronome was much harder for Spanish and Italian speakers. These results appeal to a language-particular concept of rhythmic timing. As Cummins and Port (1998) point out, “by structuring an utterance so that prominent events lie at privileged phases of a higher-level prosodic unit, rhythm is seen as an organizational principle which has its roots in the coordination of complex action and its effect in the realm of prosodic structure”. Not surprisingly, recent findings about speech rhythm and musical rhythm in several languages have revealed that there is a close and underlying relationship between the two (Patel and Daniele, 2003; Patel et al., 2006; Patel, 2008).

Finally, we want to acknowledge the potential role of factors such as speech rate and small differences in prosodic word and orthographic word counts on the main findings of the paper, namely the crosslinguistic difference in the patterns of duration across languages. Even though these differences in the number of PWs and/or orthographic words across language materials have been reported not to affect lengthening patterns of syllables in the target languages (see the null effects of PW boundaries on duration

for Catalan and Spanish in Prieto et al., 2010, and the very small effects in phrasally prominent positions in Turk and White (1999), it is not completely unambiguous that the differences found in the realization of lengthening patterns across languages are exclusively due to rhythmic patterns. Future studies are needed to shed more light on the exact influence of these factors on the observed effects.

## 5. Conclusions

The first goal of this study was to examine in detail the claimed contribution of syllable structure to the rhythmic differences between three languages belonging to different classes of rhythmic percept (English: “stress-timed”, Spanish: “syllable timed”, Catalan: “intermediate”). To this end, we used a set of materials that were controlled for their phonotactic properties. The results reported in this article showed the following: (a) as expected, vocalic rhythm indices such as %  $V$ ,  $\Delta V$ , and rPVI-V were strongly affected by syllable structure; (b) surprisingly, other indices such as nPVI-V and VarcoV were not affected by syllable structure and reflected a strong crosslinguistic differences between English on the one hand and Catalan/Spanish on the other. Thus, these results are a clear indication that despite phonotactic differences between languages, consistent and language-specific timing patterns still arise.

The results reported in this article have implications for the view that phonological properties of the target language such as syllable structure and vowel reduction can predict rhythmic behavior. Even though this view captures the positive relationship that is generally found between language phonotactics and durational properties, it does not account for two of the findings of this study: firstly, the fact that important timing differences arise between the three languages independently of syllable structure; and secondly, the fact that Catalan, which is predicted to be an intermediate language in terms of its phonological properties, patterns more clearly with “syllable-timed” languages. We have argued that Catalan vowel reduction, like the case of Singapore English vowel reduction, is a case of vowel reduction that is implemented primarily through spectral reduction, and thus has little effect on durational reduction compared to English. All in all, the results clearly suggest that the rhythmic percept is relatively independent of syllable structure composition and vowel reduction, two of the phonological properties of the language which have hitherto been claimed to be at the heart of the rhythm class distinction.

The second goal of the study was to investigate whether differences in the way languages instantiate two prosodic phenomena, namely, the durational marking of prosodic heads and edges, are correlated in any way with the potential rhythm class distinctions. Our results show that differences in syllable duration between prominent and non-prominent syllables in a language like English are much stronger than in Spanish and Catalan. Similarly, a language like English manifests a stronger durational

<sup>14</sup> Fant et al. point out that “Stress timing is not a matter of physical isochrony of interstress intervals but a perceptual dominance of heavy syllables, the succession of which is sensed as quasiperiodical. A language is sensed as syllable timed when the differences between stressed and unstressed syllables are reduced. This involves both a reduction of stress cues and a relatively greater precision and uniformity of unstressed syllables” (Fant et al. 1991a, pp. 363–364).

marking of prosodic boundaries. Following Fant, Kruckenberg and Nord (1991a,b) and Beckman (1992), we suggest that the language-specific realization of these durational prosodic phenomena might be an important ingredient in the rhythm percept across languages. In our view, though phonological factors such as syllable structure and vowel reduction do contribute to cross-linguistic differences in rhythm and tend to go hand in hand with established rhythmic percepts, there are other important language-specific properties that have strong durational effects.<sup>15</sup> As our data show, rhythmic differences could be partly attributed to cues to prosodic properties of the language like prominence and phrasing (see also Arvaniti, 2009), but other factors are likely to play a role as well.<sup>16</sup> Our data for Catalan, English, and Spanish show that perceived rhythm in these languages can be the result of both phonotactic and phrasal properties. Rhythmic organization could be understood as the perceptual result of a finely organized multisystemic durational system that differs across languages. At this point we need to investigate further what the main sources of language-specific durational variability across languages are and how these map onto the percept of rhythm.

#### **Appendix A. Materials used in the reading experiment. The number in parentheses represents the total number of syllables**

##### *A.1. Type 1. Predominantly CV-type utterances*

- Cat: La mare de la Jana és de Badalona (14)  
 Eng: The mother of Susana is from Badalona (13)  
 Span: La madre de Susana es de Badalona (14)  
 Cat: La banana de Guatemala és de bona qualitat (16)  
 Eng: The banana from Guatemala has an extra quality (15)  
 Span: La banana de Guatemala es de buena calidad (16)  
 Cat: La mare de la Jana és una bona profesora (16)  
 Eng: The mother of Susana is a very good professor (15)  
 Span: La madre de Susana es una buena profesora (16)  
 Cat: El logo de la fàbrica va ser dissenyat a Catalunya (18)  
 Eng: The company's logo was created in Catalonia (16)  
 Span: El logo de la fábrica se diseñó en Cataluña (16)  
 Cat: Canadà i Perú estan a Centreamèrica (13)  
 Eng: Canada and Peru are in Central America (14)  
 Span: Canadá y Perú están en Centroamérica (13)

<sup>15</sup> We agree with Fant et al. (1991a, p. 363) that we need to investigate the potential effect of other prosodic characteristics such as intonation patterns, voice source dynamics, and rhythmic and/or prominence perception.

<sup>16</sup> For example, there are potentially other, non-structural factors related to variation in phonetic implementation that affect rhythmic performance, as a recent study comparing child-directed and adult-directed speech for English, Catalan, and Spanish, by Payne et al. (2010), has shown.

##### *A.2. Type 2. Predominantly CVC-type utterances*

- Cat: Els mangos del Brasil i Ceilan són de qualitat extra (16)  
 Eng: Those mangos from Brazil and Ceylon are not good quality (15)  
 Span: Los mangos del Brasil y Ceilán son de calidad extra (16)  
 Cat: Els donuts d'Amsterdam són realment internacionals (15)  
 Eng: These doughnuts from Amsterdam taste almost exceptional (14)  
 Span: Los donuts de Amsterdam són realmente internacionales (17)  
 Cat: Les taronges de Londres no són les més dolces del món (15)  
 Eng: These oranges from London aren't the cleanest in the world (15)  
 Span: Las naranjas de Londres no són las más dulces del mundo (16)  
 Cat: El míting del club de tennis va ser en el pàrquing del club. (15)  
 Eng: One meeting of his golf club was in the club's parking space (14)  
 Span: El míting del club de tenis no fue en el pàrquing del club (15)  
 Cat: El doctor Frankenstein es un monstre sentimental i internacional (20)  
 Eng: Doctor Frankenstein was a big sentimental and exceptional monster (19)  
 Span: El doctor Frankenstein es un monstruo sentimental e internacional (20)  
 Type 3. Mixed utterances (coming from Ramus et al.'s corpus).

##### *A.3. English*

1. A hurricane was announced this afternoon on the TV (15)
2. The committee will meet this afternoon for a special debate (16)
3. This rugby season promises to be a very exciting one (17)
4. Having a big car is not something I would recommend in this city (18)
5. The government is planning a reform of the educational program (19)
6. My grand-parent's neighbour is the most charming person I know (15)
7. The parents quietly crossed the dark room and approached the boy's bed (16)

8. Science has acquired an important place in western society (17)
9. They didn't hear the good news until last week on their visit to their friends (18)
10. This year's Chinese delegation was not nearly as impressive as last year's (19)
11. Much more money will be needed to make this project succeed (15)
12. This supermarket had to close due to economic problems (16)
13. The last concert given at the opera was a tremendous success (17)
14. Finding a job is difficult in the present economic climate (18)
15. The city council has decided to renovate the medieval center (19)
16. The local train left the station more than five minutes ago (15)
17. In this famous coffee shop you will eat the best donoughts in town (16)
18. In this case, the easiest solution seems to appeal to the high court (17)
19. The library is opened every day from eight A.M. to six P.M (18)
20. No welcome speech will be delivered without the press officer's agreement (19)

#### A.4. Catalan

1. Ell mai va tenir la possibilitat d'expressar-se (15)
2. El lladre va fugir amb el rellotge d'or del meu pare (16)
3. Ja fa més de quinze minuts que un tren ha arribat a l'estació (17)
4. Els països occidentals no sortiran de la crisi actual (18)
5. L'art contemporani sembla tenir cada cop més bona acollida (19)
6. Una pintura de gran valor va ser robada ahir (15)
7. Hi vaig haver d'anar tan ràpid com em va ser possible (16)
8. Els veïns dels meus avis són una parella molt agradable (17)
9. Cada dia les mares surten més ràpid de la maternitat (18)
10. Aquest petit palau és un monument històric que té un gran valor (19)
11. S'assabentaren de la notícia en aquest diari (15)
12. Els pares es van acostar cap al noi sense fer soroll (16)
13. Sembla que farà molt bon temps a la costa mediterrània (17)
14. La caiguda de les taxes d'interès va ser molt apreciable (18)
15. El general va declarar que la situació estava sota control (19)

16. Van donar la notícia per ràdio dimecres passat (15)
17. Aquella botiga estarà oberta durant tot el dia (16)
18. Va dirigir el seu últim concert al teatre municipal (17)
19. Moltíssima gent va venir a celebrar la victòria amb nosaltres (18)
20. El pressupost de la conselleria de cultura va baixar molt (19)

#### A.5. Spanish

1. Se enteraron de la noticia en este diario (15)
2. Tuve que ir a buscarlo lo más rápido posible (16)
3. Los vecinos de mis abuelos son gente muy agradable (17)
4. Las madres salen cada vez mas rápido de la maternidad (18)
5. El director declaró que la situación estaba bajo control (19)
6. Hubo inundaciones graves en el verano (15)
7. La tienda está abierta durante todo el día (16)
8. Dió su último concierto en el teatro municipal (17)
9. La reconstrucción de la ciudad empezó el año pasado (18)
10. Lluve durante todo el año en los países tropicales (19)
11. El niño se levantó temprano para ver el sol (15)
12. Hubo una huelga en pleno centro de la ciudad (16)
13. Hace ya cinco minutos que el tren llegó a la estación (17)
14. Mucha gente vino a celebrar la victoria con nosotros (18)
15. No entendí nada del libro que tu me prestaste la otra vez (19)
16. Los niños salen todos los días a la misma hora (15)
17. Para eso necesitaremos mucho más dinero (16)
18. Los recientes acontecimientos hicieron escándalo (17)
19. La caída de las tasas de interés fue muy apreciable (18)
20. El presupuesto del ministerio de la cultura bajó mucho (19)

#### References

- Abercrombie, D., 1967. *Elements of General Phonetics*. Edinburgh University Press, Edinburgh.
- Arvaniti, A., 2009. Rhythm, Timing and the Timing of Rhythm. *Phonetica* 66, 46–63.
- Astruc, L., Prieto, P., 2006. Acoustic cues of stress and accent in Catalan. In: Hoffmann, R., Mixdorff, H. (Eds.), *Proc. Speech Prosody 2006*. TUDpress Verlag der Wissenschaften GmbH, Dresden, pp. 337–340.
- Astruc, L., Prieto, P., Payne, E., Post, B., Vanrell, M.M., 2009. Acquisition of tonal targets in Catalan, Spanish, and English. *Cambridge Occasional Papers in Linguistics*, vol. 5, pp. 1–14.
- Asu, E., Nolan, F., 2006. Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In: Hoffmann,



- R., Mixdorff, H. (Eds.), *Proc. Speech Prosody 2006*. TUDpress Verlag der Wissenschaften GmbH, Dresden, pp. 49–252.
- Baayen, R.H., Davidson, D.J., Bates, D., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Memory Lang.* 59, 390–412.
- Barnes, J., 2001. Domain-initial strengthening and the phonetics and phonology of positional neutralization. Paper Presented at the North East Linguistics Society (NELS), 32nd Annual Meeting, October 19–21, 2001. CUNY Graduate Center & New York University.
- Barry, W., Andreeva, B., Koreman, J., 2009. Do rhythm measures reflect perceived rhythm? *Phonetica* 66, 78–94.
- Beckman, M.E., 1992. Evidence for speech rhythms across languages. In: Tohkura, Y., Vatikiotis-Bateson, E., Sagisaka, Y. (Eds.), *Speech Perception, Production and Linguistic Structure*. IOS Press, Oxford, pp. 457–463.
- Beckman, M., Pierrehumbert, J., 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3, pp. 255–309.
- Beckman, M., Edwards, J., 1994. Articulatory evidence for differentiating stress categories, phonological structure and phonetic form. In: Keating, P.A. (Ed.), *Papers in Laboratory Phonology III*. CUP, Cambridge.
- Beckman, M., Díaz-Campos, M., McGory, J.T., Morgan, T.A., 2002. Intonation across Spanish, in the Tones and Break Indices framework. *Probus* 14, 9–36.
- Borzone de Manrique, A.M.B., Signorini, A., 1983. Segmental duration and rhythm in Spanish. *J. Phonetics* 11, 117–128.
- Boersma, P., Weenink, D. (2007). Praat: doing phonetics by computer (Version 5.1.12) [Computer program]. <http://www.praat.org/> (retrieved 04.08.09).
- Bolinger, D., 1965. *Pitch Accent and Sentence Rhythm, Forms of English: Accent, Morpheme, Order*. Harvard University Press, Cambridge, MA.
- Byrd, D., 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, 3–16.
- Carter, P.M., 2005. Quantifying rhythmic differences between Spanish, English, and Hispanic English. In: Gess, R.S., Rubin, E.J. (Eds.), *Theoretical and Experimental Approaches to Romance Linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages*, . In: *Current Issues in Linguistic Theory*, vol. 272. John Benjamins, Amsterdam, Philadelphia, pp. 63–75.
- Cummins, F., 2002. Speech rhythm and rhythmic taxonomy. In: Bel, B., Marlien, I., (Eds.), *Proc. Speech Prosody 2002*, Aix-en-Provence. pp. 121–126.
- Cummins, F., Port, R., 1998. Rhythmic constraints on stress timing in English. *J. Phonetics* 26, 145–171.
- Dauer, R.M., 1983. Stress-timing and syllable-timing reanalyzed. *J. Phonetics* 11, 51–62.
- Dauer, R.M., 1987. Phonetic and phonological components of language rhythm. In: *Proc. 11th Internat. Congress of Phonetic Sciences*, Tallinn, pp. 447–450.
- Delattre, P., 1966. A comparison of syllable length conditioning among languages. *Int. Rev. Appl. Linguist. Lang. Teach.* 4, 183–198.
- Dellwo, V., 2006. Rhythm and speech rate: a variation coefficient for delta C. In: Karnowski, P., Szigeti, I., (Eds.), *Proc. 38th Linguistic Colloquium: Language and Language Processing*, Piliscsaba 2003, Peter Lang, Frankfurt, pp. 231–241.
- Dellwo, V., Steiner, I., Aschenberner, B., Dankovicová, J., Wagner, P., 2004. The BonnTempo-Corpus and BonnTempo-Tools. A database for the combined study of speech rhythm and rate. In: *Proc. 8th ICSLP*, Jeju Island, Korea.
- Dellwo, V., Wagner, P., 2003. Relations between language rhythm and speech rate. In: Solé, M.J., Recasens, D., Romero, J., (Eds.), *Proc. 15th Internat. Congress of Phonetic Sciences*, Barcelona, pp. 471–474.
- den Os, E., 1988. *Rhythm and Tempo of Dutch and Italian: A Contrastive Study*. Elinkwijk BV, Utrecht.
- Dilley, L., Shattuck-Hufnagel, S., Ostendorf, M., 1996. Glottalization of word-initial vowels as a function of prosodic structure. *J. Phonetics* 26, 45–69.
- Estebas-Vilaplana, E., Prieto, P., 2010. Castilian Spanish intonation. In: Prieto, P., Roseano, P. (Eds.), *Transcription of Intonation of the Spanish Language*. München, Lincom Europa, pp. 17–48.
- Fant, G., Kruckenberg, A., Nord, L., 1991a. Durational correlates of stress in Swedish, French and English. *J. Phonetics* 19, 351–365.
- Fant, G., Kruckenberg, A., Nord, L., 1991b. Language specific patterns of prosodic and segmental structures in Swedish, French and English. In: *12th International Congress of the Phonetic Sciences*, Aix-en-Provence, pp. 118–121.
- Ferragne, E., Pellegrino, F., 2004. A comparative account of the supra-segmental and rhythmic features of British English dialects. In: *Proc. “Modelisations pour l’Identification des Langues”*, Paris.
- Frota, S., Vigário, M., 2001. On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus* 13, 247–275.
- Frota, S., D’Imperio, M., Elordieta, G., Prieto, P., Vigário, M., 2007. The phonetics and phonology of intonational phrasing in Romance. In: Prieto, P., Mascaró, J., Solé, M.J. (Eds.), *Segmental and Prosodic Issues in Romance Phonology*. John Benjamins, Amsterdam, Philadelphia, pp. 131–153.
- Fougeron, C., Keating, P.A., 1996. Articulatory strengthening in prosodic domain-initial position. In: *UCLA Working Papers in Phonetics*, vol. 92, pp. 61–87.
- Gavaldà-Ferré, N., 2007. *Vowel Reduction and Catalan Speech Rhythm*, Unpublished MA Thesis, University College, London.
- Grabe, E., Low, E.L., 2002. Durational variability in speech and the rhythmic class hypothesis. In: Warner, N., Gussenhoven, C. (Eds.), *Papers in Laboratory Phonology 7*. Mouton de Gruyter, Berlin, pp. 515–546.
- Hualde, J.I., 2005. *The Sounds of Spanish*. CUP, Cambridge.
- IBM Corporation, 2010. *IBM SPSS Statistics (version 19.0.0)*. Computer Program. <<http://www.spss.com/software/statistics/>>.
- Lloyd, J., 1940. *Speech Signals in Telephony*. Sir I. Pitman & Sons, Ltd., London.
- Low, E.L., Grabe, E., Nolan, F., 2000. Quantitative characterisations of speech rhythm: ‘syllable-timing’ in Singapore English. *Lang. Speech* 43, 377–401.
- Mascaró, J., 2002. Reducció vocàlica. In: Solà, J., Lloret, M.R., Mascaró, J., Pérez-Saldanya, M. (Eds.), *Gramàtica del català contemporani*. Editorial Empúries, Barcelona, pp. 89–123.
- Nespor, M., 1990. On the rhythm parameter in phonology. In: Roca, I.M. (Ed.), *Logical Issues in Language Acquisition*. Foris, Dordrecht, pp. 157–175.
- Nolan, F., Asu, E.L., 2009. The pairwise variability index and coexisting rhythms in language. *Phonetica* 66, 64–77.
- Oller, D.K., 1973. The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am.* 54, 1235–1247.
- O’Rourke, E., 2008. Speech rhythm variation in dialects of Spanish: applying the pairwise variability index and variation coefficients to Peruvian Spanish. In: *Proc. Fourth Conf. on Speech Prosody*, May 6–9, 2008, Campinas, Brazil.
- Ortega-Llebaria, M., Prieto, P., 2007. Stress and focus in Spanish and Catalan: patterns of duration and vowel quality. In: Prieto, P., Mascaró, J., Solé, M.J. (Eds.), *Segmental and Prosodic Issues in Romance Phonology*, . In: *Current Issues in Linguistic Theory Series*. John Benjamins, Amsterdam, Philadelphia, pp. 155–175.
- Ortega-Llebaria, M., Prieto, P., 2010. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Lang. Speech* 54 (1), 1–25.
- Patel, A.D., Daniele, J.R., 2003. An empirical comparison of rhythm in language and music. *Cognition* 87, B35–B45.
- Patel, A.D., Iversen, J.R., Rosenberg, J.C., 2006. Comparing the rhythm and melody of speech and music: the case of British English and French. *J. Acoust. Soc. Am.* 119, 3034–3047.
- Patel, A.D., 2008. *Music, Language, and the Brain*. Oxford University Press, Oxford.
- Payne, E., Post, B., Astruc, L., Prieto, P., Vanrell, M.M., 2009. Rhythmic modification in child directed speech. *Oxford Working Papers in Linguistics*, vol. 12, pp. 123–144.

- Payne, E., Post, B., Astruc, L., Prieto, P., Vanrell, M.M., 2010. Rhythmic modification in child directed speech. In: Russo, M., (Ed.), *Atti del convegno di Prosodia, Gli universali prosodici: confronto e ricerche sulla modellizzazione ritmica e sulle tipologie ritmiche*, Aracne Biblioteca di Linguistics, Roma, pp. 147–183.
- Payne, E., Post, B., Astruc, L., Prieto, P., Vanrell, M.M. (in press). Measuring child rhythm. *Language and Speech* 55 (2).
- Peterson, G.E., Lehiste, I., 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Am.* 32, 693–703.
- Prieto, P., Estebas-Vilaplana, E., Vanrell, M.M., 2010. The relevance of prosodic structure in tonal articulation. Edge effects at the prosodic word level in Catalan and Spanish. *Journal of Phonetics* 38 (4), 688–707.
- Prieto, P. (in press). The intonational phonology of Catalan. In: Jun, S.A., (Ed.), *Prosodic Typology 2. The Phonology of Intonation and Phrasing*. Oxford University Press, Oxford.
- Pike, K., 1945. *The Intonation of American English*. University of Michigan Press, Ann Arbor.
- Ramus, F., Nespore, M., Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292.
- Ramus, F., Dupoux, E., Mehler, J., 2003. The psychological reality of rhythm classes: perceptual studies. In: Solé, M.J., Recasens, D., Romero, J., (Eds.), *Proc. 15th International Congress of Phonetic Sciences*, Barcelona, pp. 337–342.
- Randolph, J.J., 2005. Free-marginal multirater kappa: an alternative to Fleiss' fixed-marginal multirater kappa. In: Paper presented at the Joensuu University Learning and Instruction Symposium 2005, Joensuu, Finland, October 14–15, 2005 (ERIC Document Reproduction Service No. ED490661).
- Randolph, J.J., 2008. Online Kappa Calculator. <<http://www.justus.randolph.name/kappa>> (retrieved 10.03.10).
- Shen, Y., Peterson, G.G., 1962. Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers* 9, pp. 1–36.
- Roach, P., 1982. On the distinction between “stress-timed” and “syllable-timed” languages. In: Crystal, D. (Ed.), *Linguistic Controversies*. Edward Arnold, London.
- Russo, M., Barry, W.J., 2008. Isochrony reconsidered: free-marginal kappas d. Objectifying relations between rhythm measures and speech tempo. In: *Proc. Fourth Conference on Speech Prosody*, May 6–9, 2008, Campinas, Brazil.
- Syrdal, A., & J. McGory (2000). Inter-transcriber reliability of ToBI prosodic labeling. In: *Proc. ICSLP*, pp. 235–238.
- Turk, A.E., Sawusch, J.R., 1997. The domain of accentual lengthening in American English. *J. Phonetics* 25, 25–41.
- Turk, A.E., White, L., 1999. Structural influences on accentual lengthening in English. *J. Phonetics* 27, 171–206.
- Vanrell, M.M., Prieto, P., Astruc, L., Payne, E., Post, B. (2010). Early acquisition of F0 alignment and scaling patterns in Catalan and Spanish. *Proc. Speech Prosody 2010*, (11–14 May, Chicago), 100839, pp. 1–4.
- Warrens, M.J., 2010. Inequalities between multi-rater kappas. *Adv. Data Anal. Classif.* doi:10.1007/s11634-010-0073-4.
- Wenk, B., Wiolland, F., 1982. Is French really syllable-timed? *J. Phonetics* 10, 193–216.
- White, L., Mattys, S.L., 2007a. Calibrating rhythm: first language and second language studies. *J. Phonetics* 35, 501–522.
- White, L., Mattys, S.L., 2007b. Rhythmic typology and variation in first and second languages. In: Prieto, P., Mascaró, J., Solé, M.J. (Eds.), *Segmental and Prosodic Issues in Romance Phonology*, . In: *Current Issues in Linguistic Theory Series*. John Benjamins, Amsterdam, Philadelphia, pp. 237–257.
- White, L., Payne, L., Mattys, S.L., 2009. Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In: Vigário, M., Frota, S., Freitas, M.J. (Eds.), *Phonetics and Phonology*. John Benjamins, Amsterdam, Philadelphia.
- Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717.
- Yoon, T.J., Cole, J., Hasegawa-Johnson, M., 2007. On the edge: acoustic cues to layered prosodic domains. In: Trouvain, J., Barry, W.J. (Eds.), *Proc. XVIth Internat. Congress of Phonetic Sciences*. Dudweiler, Pirrot GmbH, pp. 1264–1267.
- Yoon, T., Chavarria, S., Cole, J., Hasegawa-Johnson, M., 2004. Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In: *Proc. ICSA Internat. Conf. on Spoken Language Processing*, Jeju Island, Korea, October 4–8, 2004, pp. 2729–2732.