Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish*

Pilar Prieto[1], Maria del Mar Vanrell[2], Lluïsa Astruc[3], Elinor Payne[4], and Brechtje Post[5]
[1]ICREA-Universitat Pompeu Fabra, [2]Universitat Pompeu Fabra and Universitat Autònoma de Barcelona, [3]The Open University, [4]Oxford University, [5]University of Cambridge

Abstract

The goal of this study is twofold. First, we 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. Second, we investigate differences in the durational marking of prosodic heads and final edges of prosodic constituents between the three languages and examine 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 clearly 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, ΔV, 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.

* 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, US). We are grateful to the audience of these three conferences, and especially to M. D'Imperio, S. Frota, J.I. Hualde, F. Nolan, and L. White for very fruitful discussions on some of the issues raised in this article. 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 for performing the segmentation and coding of the data. Special thanks are due to F. Ramus et al. for sending the sentences used for 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*.

0. 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, Abercrombie 1967, among many others. This typological distinction was originally hypothesized to be instantiated through 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, it varies in direct proportion to the number of syllables these intervals contain (Shen & Peterson 1962, Bolinger 1965, Roach 1982, den Os 1988, among others). Bolinger (1965) also showed that other factors have an effect on the duration of the interstress intervals, e.g. the specific types of syllables as well as the position of the interval within the utterance. 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[1]. Against expectation, foot duration showed more variability in "stress-timed" languages than in "syllable-timed" languages. With respect to syllable duration, even 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 & Signorini (1983) showed that syllable duration in Spanish is not constant and that interstress intervals tend to cluster around an average duration. Similarly, Wenk & 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, Nespor & Mehler 1999)[2]. 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

[1] 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.

[2] Ramus, Nespor & Mehler (1999) were the first to refer to Dauer's approach with this term: "Henceforth we will refer to this position [Dauer's position] as to the phonological account of rhythm." (p. 269).

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, Nespor & Mehler’s (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 newborns could discriminate between different rhythm types. Ramus and colleagues found that three acoustic indices based on vocalic and consonantal interval measures, namely %V (or amount of vocalic stretch per utterance), ΔV and ΔC (or standard deviations in the duration of the vocalic and consonantal stretches in the utterance) were able to distinguish between perceived rhythmic classes. Ramus extended the insights of the phonological account of rhythm by extending them in the psycho-acoustic component. They pointed out that “, pointing 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, Nespor & Mehler (1999:265) 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 Measure, or PVI (Low, Grabe & Nolan 2000, Grabe & Low 2002, Asu & Nolan 2006). In addition, rate-normalised versions of Ramus’ ΔC and ΔV measures have been proposed (respectively VarcoC and VarcoV, see Dellwo 2004, and Ferragne & Pellegrino 2004). These measures are reviewed and compared for discriminatory effectiveness in White & 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 & Asu 2009 for Spanish, Frota & Vigário 2001 for European and Brazilian Portuguese, Asu & Nolan 2006 for Estonian, White & Mattys 2007a, 2007b for Dutch, French, English, and Spanish, Russo & 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. As Ramus et al. (1999:289) remarked, “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.”

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 of edges of prosodic constituents. For example, it has been shown that stressed and accented syllables are produced with additional lengthening compared with unstressed syllables (e.g. Beckman & Edwards 1994, Turk & Sawusch 1997, Turk & White 1999, among many others). Similarly, the edges of prosodic constituents have been shown to trigger lengthening effects crosslinguistically. For example, word-initial lengthening has been reported by Oller (1973), Dilley et al. (1996), Fougeron & Keating 1996, Byrd (2000), among others; and phrase-final lengthening by Wightman et al. (1992), Yoon, Cole & Hasegawa-Johnson (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 & Prieto 2007, Hualde 2005, Fant, Kruckenberg & Nord 1991a).

Recently, several authors have pointed to the need of examining 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 & 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 1992:457). Similarly, White & Mattys (2007a: 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, Payne and Mattys (2009) reinforce 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 (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 example of "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 on account 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 Catalan, having certain features distinctive of "stress-timed" languages and some other characteristics in common with "syllable-timed languages", is 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 & Low (2002) report that Catalan vocalic pairwise variability measures differ from those of Spanish, with an intermediate rhythm index. Gavaldà-Ferré (2007) undertook a comparison study between Catalan and other languages included in the Bonntempo corpus (Dellwo et al. 2004). In accordance with the previous investigations, Gavaldà-Ferré finds that data for %V and ΔC classify Catalan as a "syllable-timed" language whereas results for nPVI and rPVI characterize Catalan as an intermediate-language. The analysis of adult-directed speech carried out by Payne, Post, Astruc, Prieto & Vanrell (under revision) 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 evidence from previous work 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 the 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 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; (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 x 8 speakers x 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 greatly reduced 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 that 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 indeed 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 materials). 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 difference.

This article is organized as follows. First, the Methodology section describes the language corpus materials used for the production experiment and the methodology used for the prosodic labelling of the data. Second, the Results 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 Discussion and Conclusion sections evaluate the implications of the results for our understanding of linguistic rhythm.

1. Methodology

1.1. 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 behaviour that will be interesting 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 & 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). Even though Catalan has a greater complexity in terms of syllable structure types than Spanish (Span. *caballo*, Cat. *cavall* 'horse'; Span. *arco*, Cat. *arc* 'arch') and also in terms of proportion of closed syllables than Spanish, the predominant syllable type is CV. It also presents vowel reduction, a property which has been consistently associated with "stress-timed" languages.

Even though Catalan has been classified as an intermediate language, phonologists have not reached a firm 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: 164). This would make Catalan depart from the prototypical "syllable-timed" languages like Italian, Greek or Spanish which have a more limited range of syllable structure types and would make it approximate to "stress-timed" languages such as Dutch and English, which have a greater range of syllable structure types, including more complex types. In addition to this, Catalan undergoes a process of vowel reduction, by which the a system of seven vowels in stressed positions /i/, /e/, / ɛ/, /a/, /o/, / ɔ/, /o/, /u / are reduced to only three /i/, /u/, and [ə] 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 [ə], /o/ and /ɔ/ to /u/, and only /i/ and /u/ retain their vowel quality. Finally, Catalan, like Spanish, shows much less final lengthening than English (Ortega-Llebaria & Prieto 2007).

1.2. Materials

The experimental materials used in this investigation are falling into two 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 syllabic structure. Half of these utterances were composed of predominantly open CV-type syllables and the other half of predominantly closed syllables (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 composition (namely, number of stresses and pitch accents, and number of intended prosodic phrases). Importantly, given the morphosyntactic constituents and their length in the utterances, they were designed so that they would be pronounced in two prosodic phrases (see Frota et al. 2007 for the patterns of phrasing found across Romance languages SVO sentences). Second, we used a set of "mixed materials" that were representative of the target language and that would act as 'control' sentences in the analysis of the effects of syllable structure. For this, we used exactly the same set of sentences used by Ramus et al. (1999) for Catalan, English, and Spanish. These sentences were short news-like declarative statements initially written in French that were translated into the target language so that they were matched across languages for number of syllables (from 15 to 19).[3] (1) shows one target utterance from each language, for each of the categories (CV-type, CVC-type, and mixed). The number of syllables appears in parenthesis. The whole set of target utterances can be found in the Appendix.

[3] We would like to thank F. Ramus et al. for sending us the Spanish, English, and Catalan target written sentences used for their 1999 paper.

(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 Ámsterdam son realmente internacionales. (15)

Mixed (Ramus' target utterances)

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)

1.3. Subjects and recording procedure

A total of 24 speakers read the 30 target utterances at a normal speech rate: 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 female speakers between the ages of 28 and 40.

Recordings were performed 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 read the sentences. When errors or hesitations occurred during the readings, 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 which investigated the rhythmic and intonational properties of these languages in children's and child-directed speech (see Astruc et al. 2009, 2010; Payne et al. 2009, 2010, under revision, Vanrell et al. 2010).

The total number of utterances analyzed were 720 (24 speakers x 30 utterances x languages). The total number of syllables analyzed was 12086, and the total number of segments analyzed was 29151. All the data is available from the project website: http://april-project.info/.

1.4. Data segmentation

The segmental and prosodic labelling was performed by a research assitant, a trained phonetician who is a native speaker of English and who is also proficient in Spanish[4]. The Catalan and Spanish labellings were checked and where necessary revised by the three author native speakers of these two languages. Vocalic and consonantal segmentation were performed using *Praat* (Boersma & Weenink 2007) and according to standard criteria, e.g. F2-onset for vowels (Peterson & 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. Prevocalic glides were considered to be part of the consonantal intervals; postvocalic glides part of the vocalic intervals (e.g., the first syllable of *Guatemala* was treated as CCV; the first syllable of *Ceilán* 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.

Prosodic labelling 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 labelled with two levels of phrasing, namely the intonational phrase (or *IP*) and the intermediate phrase (or *ip*, see the original Beckman & Pierrehumbert's 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 final phrase accent. This phrasing distinction has also been proposed in the Catalan or Spanish versions of ToBI, Cat_ToBI (Prieto et al. 2009, Prieto in press) and Sp_ToBI (Beckman et al. 2002, Estebas-Vilaplana & Prieto 2010).[5] In addition to tonal marking, which has to be present at the end of both constituents, durational lengthening has also been claimed to be of a lesser intensity at the end of the former constituent (Wightman et al. 1992, Yoon, Cole & Hasegawa-Johnson 2007 for English). For our labelling, the criterion for an *IP* break was the presence of a pause. 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,

---

[4] We would like to thank Naomi Hilton for performing the segmental and prosodic coding of the data in the three languages.

[5] The reader can access both the *Cat_ToBI Training Materials* and the *Sp_ToBI Training Materials*, together with audio files and exercices, at the following web pages:
http://prosodia.upf.edu/cat_tobi/en/index.php (Cat_ToBI) and
http://prosodia.upf.edu/sp_tobi/en/index.php (Sp_ToBI).

lexically stressed, accented, and nuclear accented syllable, which have all been associated with significant lengthening (for accentual lengthening, see for example Turk & Sawusch 1997, Turk & White 1999).

Figure 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, that is, beginning of a prosodic domain (= b), end of an intermediate phrase (=e), and end of an intonational phrase (=ef), together with pause markings (= p).

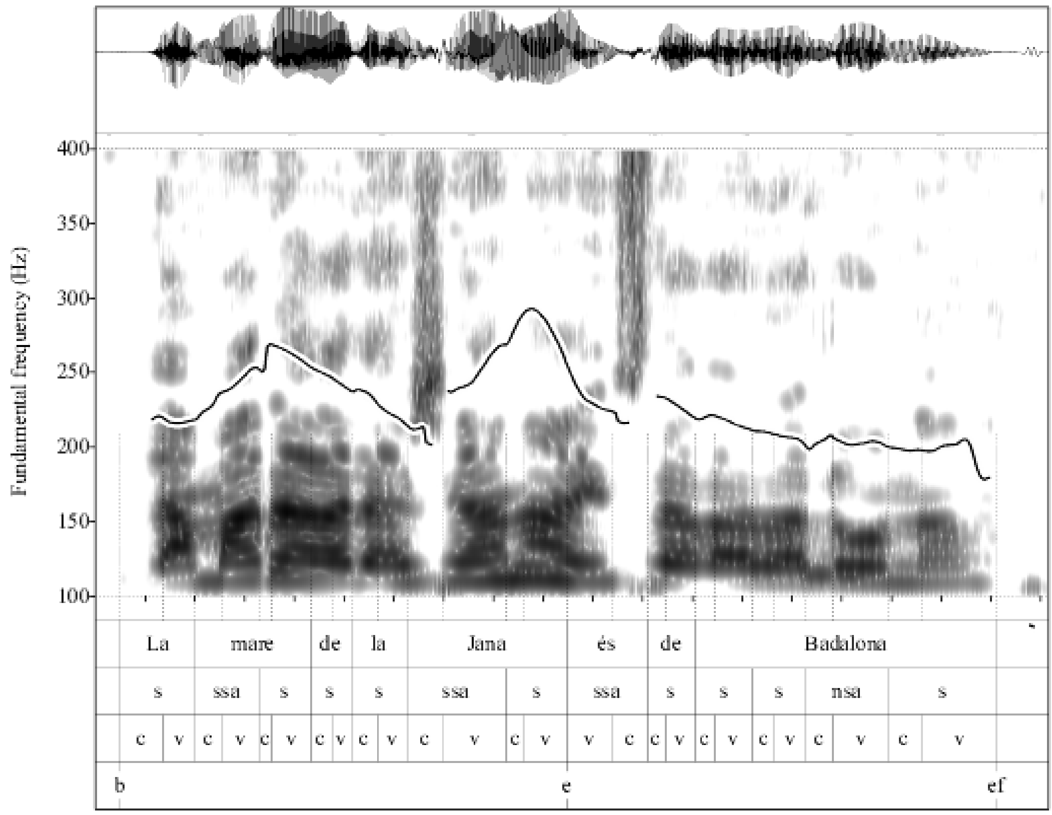


Figure 1. Waveform, spectrogram, f0 contour, and labelling schema used for the Catalan utterance *La mare de la Jana és de Badalona* ‘Jana’s mother is from Badalona’ (speaker mSMN).

1.5. Rhythm metrics

In this section we described the rhythm metrics that have been applied to the data. After data segmentation and prosodic labelling, we extracted vocalic and consonantal intervals and applied a selection of the most widely accepted rhythm metrics: %V, ΔV and ΔC following Ramus et al. (1999), nPVI-V and rPVI-V following Grabe & Low (2002), and VarcoC and VarcoV following Dellwo (2004) and Ferragne & Pellegrino (2004). In general, normalised metrics are better performing than non-normalised metrics for vocalic variability, whereas the reverse is the case for consonantal variability. And general vocalic measures discriminate better than consonantal measures (see also White & Mattys 2007a, Payne et al. under review, for a comprehensive review). Below we explain how these measures are calculated:

(1) Interval measures, which can be raw and normalized, as follows:

1a) Raw interval measures %V, ΔV, and ΔC.; Ramus et al., 1999; Ramus et al., 2003)

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);
- ΔV: the standard deviation of the duration of vocalic intervals within each sentence;
- ΔC: 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 ΔC and %V to be the most useful in discriminating across categories.

1b) Rate-normalized measures:VarcoV and VarcoC; Dellwo 2004, 2006, Ferragne & Pellegrino 2004)

Evidence that ΔV and ΔC were inversely proportional to speech rate (e.g. Dellwo & Wagner 2003), as well as criticisms by Low et al. (2000) and Grabe & Low (2002) of the standard deviation measures proposed by Ramus et al. (1999) led these authors to propose rate-normalized rhythm measures. Specifically, Dellwo (2004, 2006) proposed the normalized version of the consonantal standard deviation measure, namely, VarcoC. Later, Ferragne & Pellegrino (2004) and White & 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).

b) Pairwise variability metrics (PVIs): Grabe & Low (2002), Low, Grabe & Nolan (2000)

A different approach to measuring rhythmic differences across languages was proposed by Low, Grabe & Nolan (2000), Grabe & Low (2002), who took the sequential relationships between units in speech into account in quantifying variability in vocalic and consonantal intervals. The PVI measures express the average difference between adjacent units such as vowels, consonantal intervals, syllables, or feet (see also Asu & Nolan 2006).[6] 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.

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

1.7. Statistical Analyses

In order to test the effects of syllable structure and language on the rhythmic correlates, we carried out 3x3 factorial ANOVAs on each of the rhythmic metrics (namely the interval measures %V, ΔV, ΔC, the Varco measures VarcoV and VarcoC, and the PVI measures for vowels nPVI and rPVI). The dependent variables were each of the rhythmic scores and the two independent variables under consideration were Syllable Type (three levels: open syllables, closed syllables, mixed) and Language (Catalan, English, Spanish). When there was an interaction between the two factors, we performed post hoc tests on the data to test differences in behaviour across languages and across syllable types. To test the effects of prominence level and phrasal position on syllable timing patterns, we carried out two 3x3 factorial ANOVAs. In the two cases, the response variable was syllable duration and the two independent variables were Prominence Level (3 levels: nuclear stressed, stressed, unstressed) 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.

[6] Low, Grabe & Nolan (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 as vowel durations were significantly more variable in British English than in Singapore English.

2. 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. Box plots 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.

2.1. The effects of syllable structure on the rhythm metrics

2.1.1. Interval measures: %V, ΔV, ΔC

As mentioned before, 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. 1999:289), one would expect this measure, together with the standard deviation for vocalic and consonantal intervals, also 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 box plot in Figure 2 shows the mean group results for %V for Catalan (white boxes), English (striped boxes), and Spanish (grey boxes). 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 structure, the higher %V is. That is, while the materials composed of predominantly open syllables range from 47.77% (English) to 55.57% (Catalan), the materials composed 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 behaviour 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 dependant on the phonotactic properties of the speech materials, crosslinguistic differences can also be captured by this metric when syllable structure is controlled for.

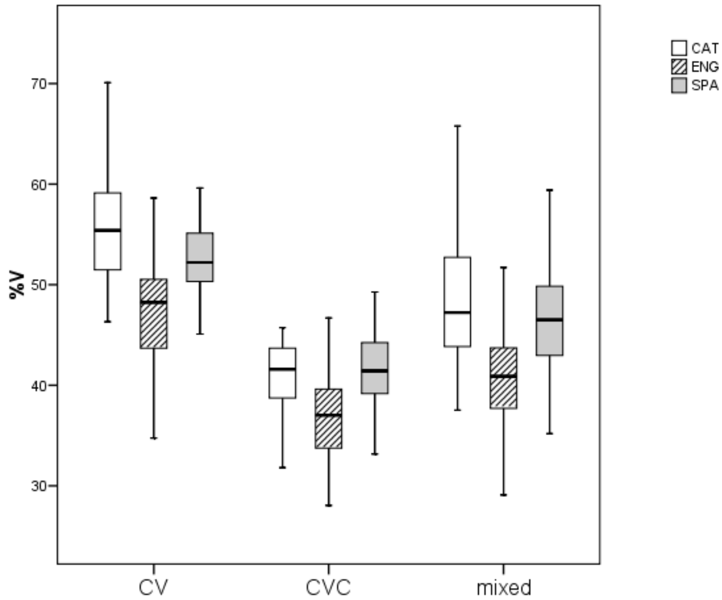


Figure 2. Box plot comparing the results of %V as a function of the phonotactic factor, for Catalan (white boxes), English (striped boxes), and Spanish (grey boxes). The x axis shows phonotactic type, that is, predominantly CV-type utterances (left), predominantly CVC-type utterances (middle), and mixed utterances (right).

A 3x3 factorial ANOVA was applied to the %V. Two main factors were taken into account, namely Syllable Type (3 levels: open, closed, mixed syllables), and Language (3 levels: Catalan, Spanish, and English). The analysis revealed a significant mean effect of Language ($F(1,2) = 65.871$, $p < 0.001$) and Syllable Type ($F(1,2) = 131.183$, $p < 0.001$) on %V, and no significant interaction between both factors, ($F(1,4) = 1.021$, $p = 0.396$). Thus, our %V results show that, as expected, even 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, 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 the phonological control and 'representativeness' of the materials is crucial when we want to compare languages using this index. We discuss this issue later in the Discussion section.

Let us now analyze the behaviour of the other two interval measures proposed by Ramus et al. (1999), namely ΔV and ΔC, or standard deviation of V and C. The two boxplots in Figure 3 show the mean of the standard deviation of the vocalic intervals, ΔV, and the mean of the standard deviation of consonantal intervals, ΔC, for all three languages. Catalan is represented by the white boxes, English by the striped boxes, and Spanish by the grey boxes. 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 distinct behaviour between ΔV and ΔC. ΔV 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, ΔC values only distinguish between these two groups of languages in the 'mixed' condition.

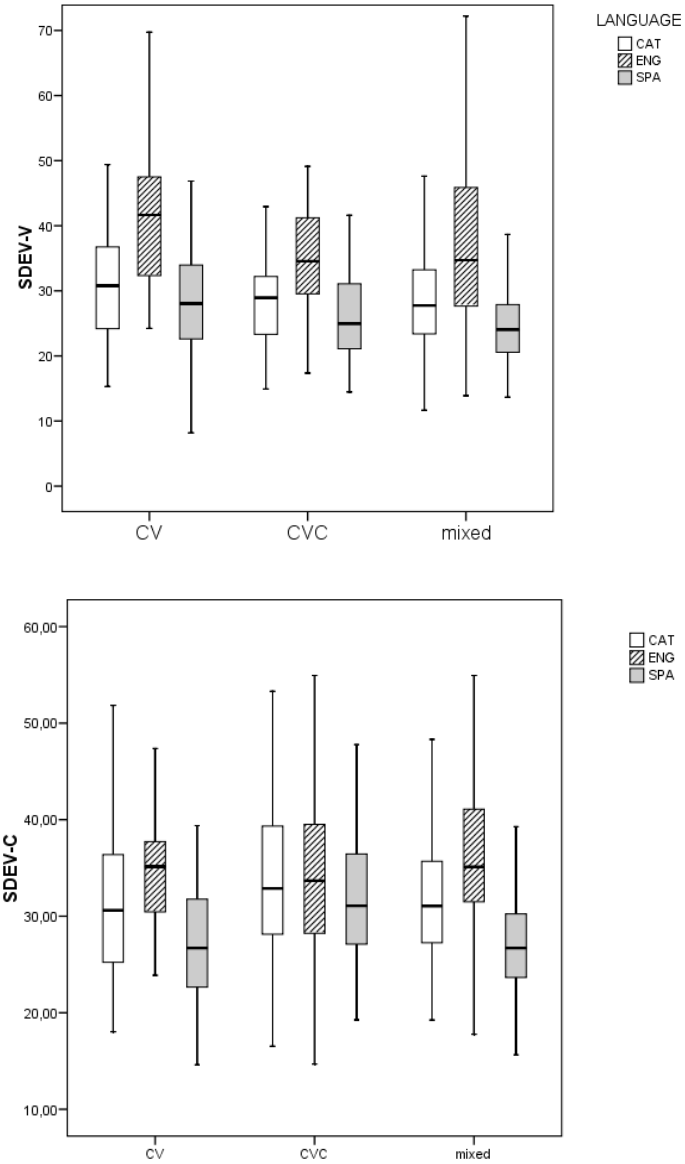


Figure 3. Mean group results for ΔV (upper graph) and ΔC (bottom graph) as a function of the phonotactic factor, for all three languages. The x axis shows phonotactic type, that is, predominantly CV-type utterances (left), predominantly CVC-type utterances (middle), and mixed utterances (right).

For ΔV, the results of a 3x3 factorial ANOVA showed main effects of Language ($F(1,2) = 84.902$; $p < 0.001$) and Syllable Type ($F(1,2) = 7.818$; $p < 0.001$), and no interaction between the two ($F(1,4) = 0.947$; $p = 0.436$). However, for ΔC we found a main effect of Language ($F(1,2) = 36.058$; $p < 0.001$), but no significant effect of Syllable Type $F(1,2) = 3.163$; $p = 0.043$); and a significant interaction between Language and Syllable

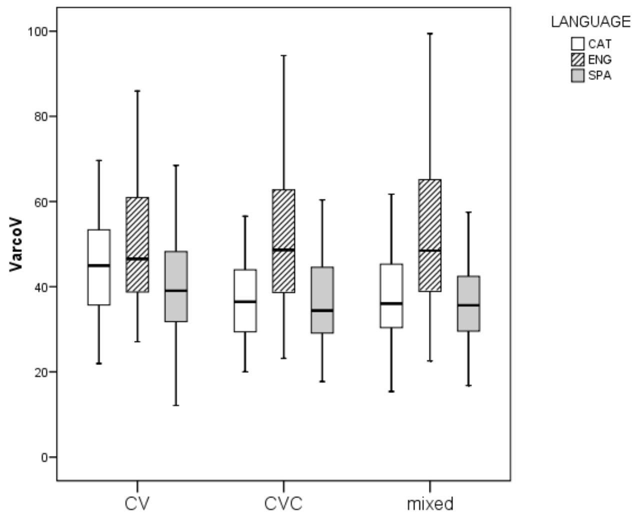
Type ($F(1,4) = 3.632$; $p < 0.007$). Planned post-hoc comparisons on the ΔC data revealed a significant difference between the CV and the mixed condition (at $p < 0.001$) and between the CV-CVC condition (at $p < .05$). Thus ΔC only discriminates between languages in the 'mixed condition', but crucially does not distinguish between languages in materials controlled for syllable structure (CV and CVC materials). This is a clear reflection of the fact that ΔC values are strongly influenced by the number of adjacent consonants found in the target utterances, and thus on the phonotactic properties of the speech materials.

Finally, the results in this section constitute a preliminary indication that differences in vocalic interval durations across languages cannot solely be the reflection of the phonotactic differences across the languages involved and that the systematic differences across %V and ΔV values across different types of materials may be reflecting a systematic crosslinguistic difference in timing organization.

2.1.2. VarcoV and VarcoC

As mentioned before, the criticisms on the raw standard deviation measures ΔC and ΔC as being speech rate dependent (Low et al. 2000, Grabe & Low 2002, Dellwo & Wagner 2003) led these authors to propose rhythmic measures that tried to correct for this dependency. Specifically, Dellwo (2004, 2006) proposed the VarcoC measure, and Ferragne & Pellegrino (2004) and White & Mattys (2007a,b) the VarcoV measure. As White and Mattys (2007a) note, even though the normalized VarcoV metric performed better at distinguishing rhythmic categories than its non-normalized counterpart, namely ΔV, this is not the case for VarcoC, which appeared to eliminate linguistically-interesting variation.

The mean group results for the calculation of VarcoV and VarcoC appear in the two boxplots in Figure 4, for Catalan (white boxes), English (striped boxes), and Spanish (grey boxes). 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 show a contrasting behaviour between VarcoV and VarcoC that clearly resembles the contrasts reported between ΔV and ΔC 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.



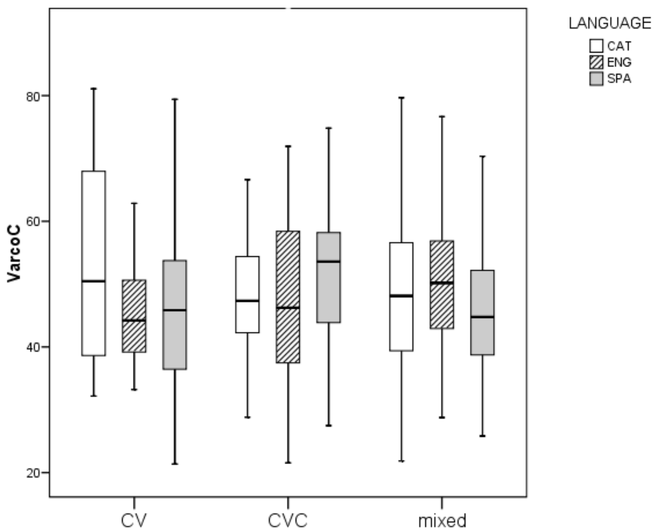


Figure 4. Mean group results for VarcoV (upper graph) and VarcoC (bottom graph) for all three languages. The x axis separates the data into predominantly CV-type utterances (left), predominantly CVC-type utterances (middle), and mixed utterances (right).

A 3x3 factorial ANOVA was applied to the VarcoV and VarcoC data. The response variables were VarcoV/VarcoC and the main factors were Syllable Type (3 levels: open vs. closed vs. mixed syllables), and Language (3 levels: Catalan, Spanish, and English). The VarcoV analysis revealed a significant main effect of Language, $F(1,2) = 49.69$; $p < 0.001$, but no effects of Syllable Type $F(1,2) = 1.312$; $p = 0.270$, nor any significant interaction between Language and Syllable Type, $F(1,4) = 1.415$; $p = 0.227$). By contrast, the VarcoC results revealed no effects of Language $F(1,2) = 1.968$; $p = 0.141$ nor of Syllable Type, $F(1,2) = 0.723$; $p = 0.486$, and a significant interaction between the two factors $F(1,4) = 5.496$; $p < 0.001$. Planned post-hoc 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 $p < 0.05$).

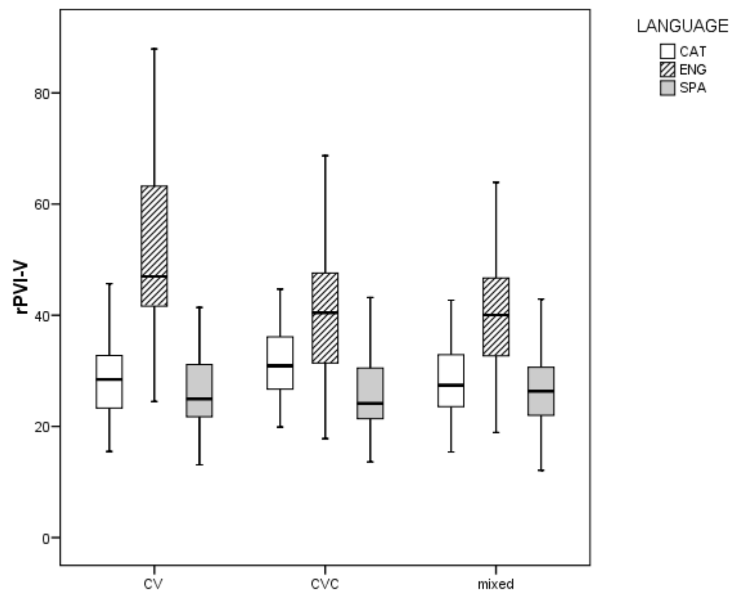
Clearly, while VarcoV successfully discriminates between English vs. Catalan and Spanish in all the materials, this is not the case with VarcoC.

Thus, the findings in this section revealed that VarcoV, in contrast with VarcoC, has a high discriminating power which is not dependent on syllable structure, and thus it can be considered to be a robust and reliable rhythm measure across languages. Moreover, VarcoV and VarcoC mirror the results reported in the previous section for the raw vowel and consonantal standard deviation data, ΔV and ΔC.

2.1.3. Pairwise Variability Indices (nPVI-V, rPVI-V)

In this section we explore the behaviour of the PVI measures in trying to capture the rhythmic differences across language materials that are controlled for syllable types. As explained in section 1.5, this measure tries to capture the degree of durational variability between adjacent units, measured sequentially (Low, Grabe & Nolan 2000, Grabe & Low 2002). In the languages under investigation, we expected to find a higher vocalic PVI value in English, reflecting more durational variability between successive vowels. Although Low, Grabe & Nolan (2000) also proposed a consonantal PVI (PVI-C), they also 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, Grabe & Nolan 2000: 396-7).

The two boxplots in Figure 5 show the group mean results of the raw vocalic measures (rPVI-V, upper graph), and the normalized measures (nPVI-V, lower graph) for the three languages. In the graphs, the data are separated into the three types of materials used, namely predominantly CV-type utterances (left), predominantly CVC-type utterances (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 and the normalized PVI measures show much effect of this factor. The PVI-V data thus mirror the behaviour of ΔV and VarcoV measures in the preceding sections, as they are relatively independent of syllable structure and do reflect stable durational differences across languages.



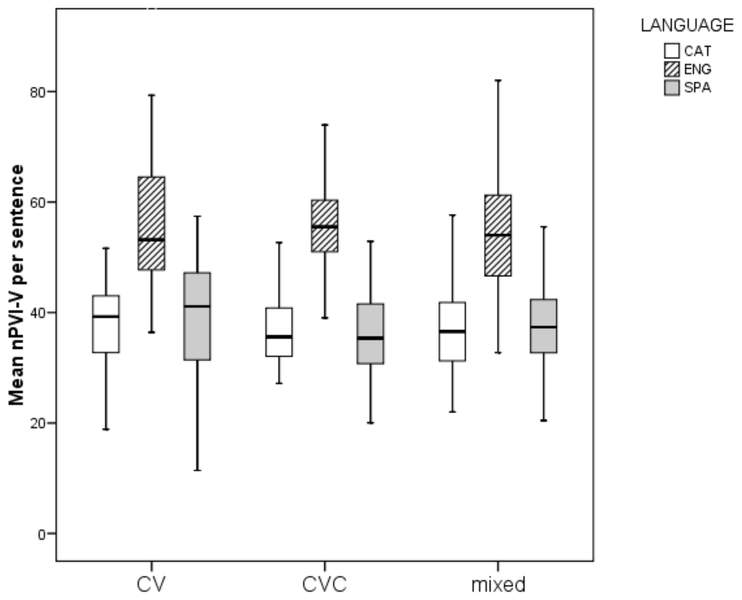


Figure 5. Group mean results of the raw vocalic Pairwise Variability Index calculations (rPVI-V, upper graph), and the normalized PVI-V measures for the three languages. The x axis separates the data into predominantly CV-type utterances (left), predominantly CVC-type utterances (middle), and mixed utterances (right).

As expected, the results of a factorial analysis of variance revealed a distinct behaviour between the rPVI-V and the nPVI-V measures. One the one hand, the raw PVI data revealed a significant main effect of both Language, $F(1,2) = 187.005$; $p < 0.001$ and Syllable Type $F(1,2) = 5.775$; $p < 0.004$ (and a significant interaction between Language and Syllable Type, $F(1,4) = 10.277$; $p < 0.001$). Planned post-hoc comparisons on the rPVI-V data showed significant differences between the CV and Mixed factors of the Syllable Type condition (at $p < 0.001$), and a significant difference between the English-Spanish and English-Catalan language pairs (at $p < 0.001$). On the other hand, the normalized PVI analysis showed a significant main effect of Language, $F(1,2) = 203.643$; $p < 0.001$ but no effects of Syllable Type $F(1,2) = 1.224$; $p = 0.295$ on nPVI-

V, and no significant interaction between Language and Syllable Type, $F(1,4) = 0.232$; $p = 0.921$), indicating that this measure is more robust in capturing rhythmic distinctions across phonotactic conditions.

To summarize, our results thus far show that different vocalic measurements, namely %V, ΔV, VarcoV, and nPVI-V provide us with a more robust index of rhythmic classes than consonantal measures. Measures of consonantal interval variability such as ΔC and VarcoC are more sensitive to phonotactic differences to the rhythmic differences. This constitutes a preliminary indication that vocalic measures constitute a more stable rhythm 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 ΔV, 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.

2.2. Effects of prominence level and phrasal position on timing patterns

In the next two subsections, we will describe the effects of phrasal prominence and phrasal position on the timing patterns found in our data.

2.2.1 Effects of prominence levels on syllable duration

In our data, phrasal prominence was labelled using the following four levels: unstressed, lexically stressed but unaccented, accented, and nuclear accented. We hypothesized that different levels of prominence will reflected in durational differences in our data, with stronger effects for English than for Catalan and Spanish (Ortega-Llebaria & Prieto, 2007 for Spanish, Astruc & Prieto 2006 for Catalan, and Turk and White 1999 for English). Since we had virtually no examples of stressed unaccented syllables (possibly due to the fact that the data was elicited through a reading task), we conflated stressed unaccented and stressed accented categories under the category “stressed”.

Figure 6 shows the mean syllable duration (in ms.) for the three languages (Catalan = white boxes, English = striped boxes, Spanish = grey boxes). The data are separated into stressed/accented positions (left), nuclear stressed position (middle), and unstressed positions (right). The graphs reveal regular and consistent patterns of behaviour across languages and across conditions. As expected, the 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 stressed/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 (Catalan 152 ms, English 158 ms, and Spanish 145 ms) indicate that there is in fact no evidence for compression of unstressed syllables in English.

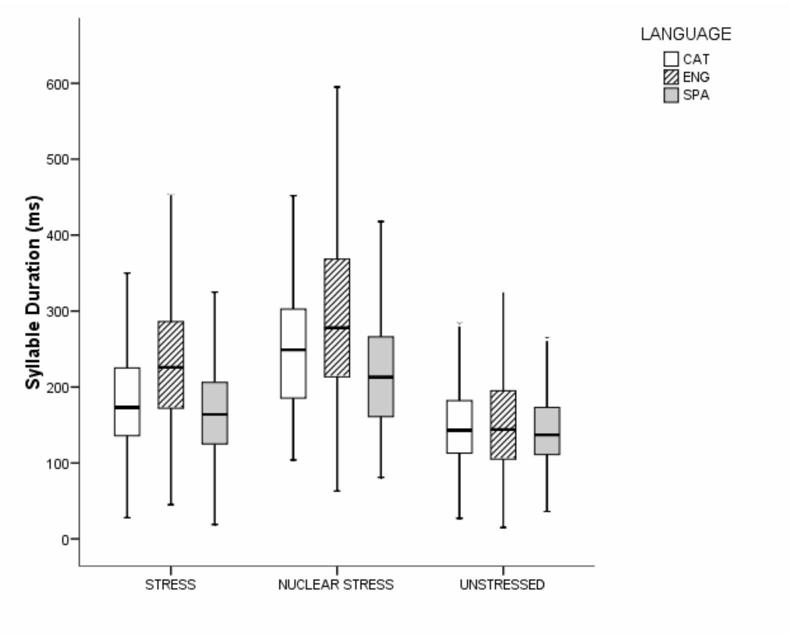


Figure 6. Mean syllable duration (in ms) in the three languages (Catalan = white boxes, English = striped boxes, Spanish = grey boxes). The data are separated into stressed (accented) positions (left), nuclear stressed position (middle), and unstressed positions (right).

A 3x3 factorial ANOVA was applied to the data. The response variable was syllable duration (in milliseconds), and the main factors were Prominence Level (3 levels: nuclear stressed, stressed, unstressed) and Language (3 levels: Catalan, English, and Spanish). Results revealed a significant main effect of Language ($F(1,2) = 71.989$; $p < 0.001$) and Prominence Level ($F(1,3) = 762.658$; $p < 0.001$) on syllable duration ($p < 0.001$) and a significant interaction between Language and Prominence Level ($F(1,4) = 39.325$; $p < 0.001$)]. Planned post-hoc comparisons reveal that the three languages differ in their durational patterns (at $p < 0.001$) and that different prominence levels are significantly different (at $p < 0.001$).

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 2005:273)[7].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 be probably related to the presence of vowel reduction, as claimed by Ortega-Llebaria & Prieto (in press). They report 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 they also have the effect of amplifying the duration differences that were triggered by stress. Interestingly, the same pattern of results was found by Fant, Kruckenberg & Nord's (1991a: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 of 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 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.

2.2.2. 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, Yoon, Cole & Hasegawa-Johnson 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.

Figure 7 shows the mean syllable duration (in ms) in the three languages (Catalan = white boxes, English = striped boxes, Spanish = grey boxes). 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 have a lengthening effect both at the level of the intermediate phrase and at the level of the intonational phrase. The results also confirm what has been claimed for English, namely, that this language has very long syllables at the end of prosodic domains. English syllables are consistently longer than

[7] As Hualde (2005: 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."

Spanish or Catalan both at the right edge of an ip (English 298 ms vs Catalan 246 ms and Spanish 250 ms) and at the right edge of an IP (English 328 ms vs. Catalan 265 ms and Spanish 224 ms). On the other hand, the differences between Catalan and Spanish are smaller. As for the potential differences between the boundary and the lengthening effect, the English data show strong differences between duration patterns at edges of an ip and of an IP. Yet this pattern is not clear neither for Catalan nor Spanish.

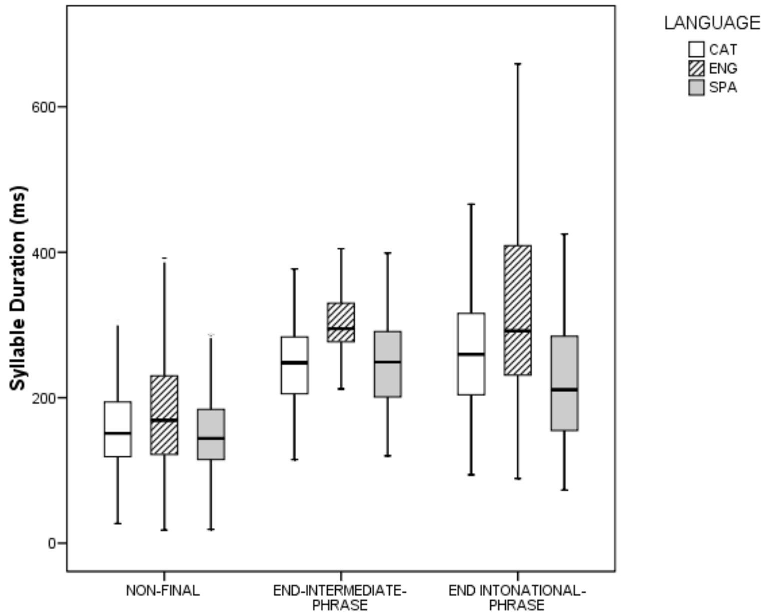


Figure 7. Mean syllable duration (in ms) in the three languages (Catalan = white boxes, English = striped boxes, Spanish = grey boxes). The data are separated into the following phrasal positions: non-final (left), end of intermediate phrase (middle), and end of intonational phrase (right).

A 3x3 factorial ANOVA was applied to the data. The response variable was syllable duration (in milliseconds), and the main factors were Phrasal Position (3 levels: non-final, end intermediate phrase, end intonational phrase) and Language (3 levels: Catalan, English, and Spanish). Results revealed a significant main effect of Language ($F(1,2) = 31.991$; $p < 0.001$) and Phrasal Position on syllable duration ($F(1,2) = 843.701$; $p < 0.001$) and no significant interaction between Language and Phrasal Position ($F(1,4) = 1.914$; $p = 0.105$.

The data in this section has revealed language-specific differences in the implementation patterns of final lengthening across the three languages. English has a stronger pre-boundary lengthening effect than Spanish and Catalan. Catalan also seems to have stronger marking of prosodic boundaries. Findings also provide strong evidence for a two-way phonological distinction in phrasing at least for English, namely the ip and the IP, confirming recent results by Yoon, Cole & Hasegawa-Johnson (2007) for American English. By contrast, this is not the case for Catalan and Spanish, confirming recent results by Vizcaino Ortega, Estebas-Vilaplana, Cabrera Abreu & Astruc (2008) for Spanish. 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. Discussion

a. *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, ΔV, ΔC, VarcoC, and rPVI. This result did not come as a surprise; since many authors had 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 such structure. As Ramus et al. (1999: 273-4) pointed out "ΔC 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 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 normalised measures of vocalic variability, VarcoV and nPVI-V —that is, the normalised measures of vocalic variability[8]-, 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, ΔV, 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 is the following. 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 ΔC 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, ΔV, and rPVI)..

Importantly, there is a consistent difference between the behaviour of vocalic and consonantal measures of rhythm. Crucially, the consonant interval metrics investigated, whether normalised or non-normalised (ΔC and VarcoC), were not able to discriminate across languages (except for ΔC in the mixed condition). The low discriminatory power of the consonantal indices indicates that these measures are highly dependant on phonotactic factors, which is just one factor contributing to the percept of rhythm. This finding is compatible with previous reports on consonantal metrics. Not surprisingly, White & 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

[8] Remember that VarcoV and nPVI-V are the non-normalised counterparts ΔV and the raw PVI-V respectively.

first and second language learners. Yet measures of consonantal interval variation (ΔC and VarcoC) proved to be far less effective or consistent in discriminating between rhythmic types. The high discriminatory power of vocalic measures (both interval measures and standard deviation measures) in comparison would make sense because the vowel portion of the string is known to better reflect the timing patterns in language (Low, Grabe & Nolan 2000, Grabe & Low 2002, among others), and therefore be more sensitive to differences in these patterns. Moreover, the high informational load encoded by vowels is an essential part of Mehler et al.'s (1996) hypothesis (see also Ramus et al. 1999), that children's early perception of speech is primarily centred on vowels.

The results in this article 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 Grabe & Nolan 2000 for the Singapore and British varieties of English, Frota & Vigário 2001 for the European and Brazilian varieties of Portuguese, O'Rouke 2008 for the Lima and Cuzco varieties of Spanish, Nolan & Asu 2009 for Peninsular Spanish and Mexican Spanish, White, Payne & Mattys 2009 for northern and southern varieties of Italian, among others). Even 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, Grabe & Nolan (2000) report a robust acoustic difference between the British English ("stress-timed") and the Singapore English ("syllable-timed") varieties, regardless of the fact that syllable structure and vowel reduction properties are similar across the two varieties. Similarly, O'Rourke 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, Post, Astruc, Prieto & Vanrell, 2009, 2010, on a comparison of the rhythm of child-directed and adult-directed speech for English, Catalan and Spanish.)

As we will see in the next subsection, 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 Nespor 1990, among others). As mentioned before, 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 results) clearly cluster Catalan with "syllable-timed" languages like Spanish. This is a clear indication that vowel reduction in this language is not correlated with durational reduction in this language, as has been argued by previous investigations (Ortega-Llebaria & Prieto in press). As we will see below, the case of Catalan suggests that durational variability cannot always be traced back to phonological properties of the language.

b. The status of Catalan

Even though Catalan has been classified as an intermediate language with respect to rhythm, linguists have not reached a firm consensus on its status. In our data, rhythm metrics do not behave consistently with respect to the classification of Catalan. Even 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, but yet significant differences between Catalan and Spanish (namely %V and ΔV). Similarly, the analysis of adult-directed speech carried out by Payne, Post, Astruc, Prieto & Vanrell (under revision) 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 not for all the rest of metrics. Gavaldà-Ferré (2007) undertook a comparison study between Catalan and other languages included in the Bonntempo corpus (Dellwo et al. 2004). She found that data for %V and ΔC 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 in no complete agreement as far as Catalan is concerned, our quantitative results taken together tend to agree with Ramus et al. (1999) in that Catalan tends to group quantitatively with the rest of "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. Even 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 are significantly different from those of Spanish. Ramus et al. (2003) already stated that vowel reduction in Catalan is not enough to make it separate from "syllable-timed" languages. They reach the following conclusion: "ΔV 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. 2003:5). In accordance with this, we claim that the "syllable-timed" behaviour of Catalan is rooted in the way Catalan implements vowel reduction, which is very different from English. In a recent study about the correlates of stress distinctions in Catalan and Spanish, Ortega-Llebaria & Prieto (in press) show that even though 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 primarily with vowel quality changes. A similar vowel reduction pattern was found in Singapore English. Low, Grabe & Nolan (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 Gavaldà-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 different types of vowel reduction: one that compresses syllable duration very strongly, such as the one 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 indeed runs against the idea that vowel reduction processes should automatically lead to a percept of "stress-timed" rhythm.

Finally, let us turn back to the issue of the classification of the Catalan language in the hypothesised rhythmic scale. As mentioned before, 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, Grabe & Nolan 2000, Gavaldà-Ferré 2007, Payne, Post, Astruc, Prieto & Vanrell under revision). Even though the different performances 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 call for 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 distinction of languages into rhythmic classes is based on an auditory impression of languages being more or less regularly timed and there are almost no perceptual studies correlating acoustic cues with rhythmic impressions.[9] 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 before, rhythmically mixed languages have been considered to be key evidence in the debate between the existence of either categorically distinct rhythmic classes or a rhythmic continuum. While the research reported by Ramus, Nespor & Mehler (1999), Low, Grabe & Nolan (2000), Grabe & Low (2002), using different quantitative measures, has shown that it is possible to

[9] We would like to thank an anonymous reviewer for pointing this out to us.

achieve useful scalar characterisation of the rhythm of different languages, it is still an open question whether we need to acknowledge a categorical rhythmic distinction between languages. If we use acoustic correlates of timing and durational variability, it might well be that differences across languages will be of a continuous nature.[10] Yet the rhythm percept of these production differences might well be categorical in nature. Thus we need to investigate more the mapping between perceptual properties of rhythm and the acoustic properties present in speech (see Barry et al. 2009 for a recent study).

c. *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 has revealed the following: (a) English has a stronger durational marking of accentual prominence than Catalan and Spanish; (b) English has 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 before, this view is not new: (see in particular Fant, Kruckenberg & Nord (1991a, 1991b) and Beckman (1992). Fant, Kruckenberg & Nord (1991a), in their study of the phonetics of stress in Swedish, English, and French claimed that the durational correlates of prominence could be a reflection of the rhythmic classification.[11] 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 its potential relationship 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 behaviour of the rhythmic indices can be traced back to the tendency of Catalan to have slightly stronger durational marking of prosodic prominence and prosodic boundaries than Spanish.

[10] 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 gradual distinction between dialects of the same language, which might in some cases reveal a change in progress.

[11] 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, Kruckenberg & Nord 1991a:363-364).

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 by Cummins (2002) have shown that while English speakers could 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 & 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 & Daniele 2003; Patel, Iversen, & Rosenberg, 2006; Patel 2008).

4. Conclusions

The first goal of this study has been to examine in detail the claimed contribution of syllable structure on the rhythmic differences between three languages belonging to different classes of rhythmic percept (English: "stress-timed", Spanish: "syllable timed", Catalan: "intermediate"). For this, 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, ΔV 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 behaviour. 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; secondly, the fact that Catalan, which is predicted to be an intermediate language from their 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 though 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 been claimed to be at the heart of the rhythm class distinction.

The second goal of this 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 marking of prosodic boundaries. Following Fant, Kruckenberg & Port (1991) 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, even though phonological factors such as syllable structure and vowel reduction 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.[12] As our data show, rhythmic differences could be partly attributed to cues to prosodic properties of the language like prominence and phrasing (cf. also Arvaniti 2009), but other factors are likely to play a role as well[13]. 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 are the main sources of language-specific durational variability across languages and how these map onto the percept of rhythm.

5. References

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66*, 46-63.

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* 5, 1-14.

Astruc, L., & Prieto, P. (2006). Acoustic cues of stress and accent in Catalan. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006* (pp. 337-340). Dresden: TUDpress Verlag der Wissenschaften GmbH.

Asu, E., & Nolan, F. (2006). Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In R. Hoffmann, & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006* (pp. 249-252). Dresden: TUDpress Verlag der Wissenschaften GmbH.

Auer, P. (1991). “Stress-Timing” vs. “Syllable-Timing” from a Typological Point of View. In B. Palek, & P. Janota (Eds.), *Proceedings of LP’90* (pp. 292–305). Prague: Charles University Press.

[12] We agree with Fant, Kruckenberg & Nord (1991a:363) that we need to investigate the potential effect of other prosodic characteristics such as intonation patterns, voice source dynamics on rhythmic and/or prominence perception.

[13] For example, there are potentially other, non-structural factors related to variation in phonetic implementation that affect the 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.

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*.

Barry, W., Andreeva, B. and Koreman, J. (2009). Do Rhythm Measures Reflect Perceived Rhythm? *Phonetica*, *66*, 78-94.

Beckman, M., & Pierrehumbert, P. (1986). *Intonational Structure in Japanese and English*. Phonology Yearbook III (pp. 15-70).

Beckman, M. E. (1992). Evidence for speech rhythms across languages. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure* (pp. 457-463). Oxford: IOS Press.

Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 68–86). Cambridge: CUP.

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories, Phonological Structure and Phonetic Form. In P. A. Keating (Ed.), *Papers in Laboratory Phonology III* (pp. 7-33). Cambridge: CUP.

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.

Bertinetto, P. M. (1981). *Strutture prosodiche dell'italiano*. Firenze: Accademia Della Crusca.

Borzone de Manrique, A. M. B., & Signorini, A. (1983). Segmental duration and rhythm in Spanish. *Journal of Phonetics, 11*, 117–128.

Boersma, P., & Weenink, D. (2007). *Praat: doing phonetics by computer* (Version 5.1.12) [Computer program]. Retrieved August 4, 2009, from http://www.praat.org/

Bolinger, D. (1965). *Pitch accent and sentence rhythm, Forms of English: Accent, morpheme, order.* Cambridge, MA: Harvard University Press.

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 R. S. Gess, & E. J. Rubin (Eds.), *Theoretical and Experimental Approaches to Romance Linguistics*: *Selected papers from the 34th Linguistic Symposium on Romance Languages (Current issues in linguistic theory 272)* (pp. 63–75). Amsterdam, Philadelphia: John Benjamins.

Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002* (pp. 121-126). Aix-en-Provence.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145-171.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*, 385-400.

Dasher, R. M., & Bolinger, D. (1982). On pre-accentual lengthening, *Journal of the International Phonetic Association*, *12*, 58-69.

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

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th International Congress of Phonetic Sciences* (pp. 447-450). Talinn.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching*, *4*, 183-198.

Dellwo, V. (2004). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski, & I. Szigeti (Eds.), *Language and Language Processing* (Proceedings of the 38th Linguistics Colloquium) (pp. 231-241). Frankfurt: Peter Lang.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski, & I. Szigeti (Eds.), *Language and language processing: Proceedings of the 38th linguistic colloquium* (pp. 231–241). Piliscsaba 2003. Frankfurt: Peter Lang.

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 471-474). Barcelona.

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. *Proceedings of the 8th ICSLP*, Jeju Island, Korea.

Den Os, E. (1988). *Rhythm and tempo of Dutch and Italian; a contrastive study*. Utrecht: Elinkwijk BV.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 26*, 45-69.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In Prieto, P. & Roseano, P. (coords.)(2010). *Transcription of Intonation of the Spanish Language*. Lincom Europa: München..

Fant, G., Kruckenberg, A., & Nord, L. (1991a). Durational correlates of stress in Swedish, French and English. *Journal of Phonetics, 19*, 351-365.

Fant, G., Kruckenberg, A., & Nord, L. (1991b). Language specific patterns of prosodic and segmental structures in Swedish, French and English. *12th International Congress of the Phonetic Sciences* (p. 118-121). Aix-en-Provence.

Ferragne, E., & Pellegrino, F. (2004). A comparative account of the suprasegmental and rhythmic features of British English dialects. In *Proceedings of "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. Segmental and Prosodic Issues in Romance Phonology, ed. by P. Prieto, J. Mascaró, and M.-J. Solé, pp. 131-153. John Benjamins: Amsterdam/Philadelphia.

Fougeron, C., & Keating, P. A. (1996). Articulatory strengthening in prosodic domain-initial position. *UCLA Working Papers in Phonetics*, *92*, 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 rhythm class hypothesis. In N. Warner, & C. Gussenhoven (Eds.), *Papers in Laboratory Phonology* 7 (pp. 515–546). Berlin: Mouton de Gruyter.

Herrick, D. (2003). *An Acoustic Analysis of Phonological Vowel Reduction in Six Varieties of Catalan*. PhD. Dissertation, University of Santa Cruz.

Hualde, J. I. (2005). *The Sounds of Spanish*. Cambridge: CUP.

Lloyd, J. (1940). *Speech Signals in Telephony*. London: Sir I. Pitman & sons, Ltd.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English. *Language and Speech*, *43*, 377-401.

Mascaró, J. (2002). Reducció vocàlica. In J. Solà, M. R. Lloret, J. Mascaró, & M. Pérez-Saldanya (Eds)., *Gramàtica del català contemporani* (pp. 89-123). Barcelona: Editorial Empúries.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant`s viewpoint. In J. L. Morgan & K. Demuth, *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 101-116). Mahwah, NJ: Lawrence Erlbaum Associates.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756-766.

Nespor, M. (1990). On the Rhythm Parameter in Phonology. In I. M. Roca (Ed.), *Logical Issues in Language Acquisition*, (pp. 157-175). Dordrecht: Foris.

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. *Journal of the Acoustical Society of America*, *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. *Proceedings of the Fourth Conference 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 P. Prieto, J. Mascaró, & M. J. Solé (Eds.), *Segmental and Prosodic Issues in Romance Phonology (Current issues in linguistic theory series)* (pp. 155-175). Amsterdam, Philadelphia: John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (in press, to appear Jan 2011). "Acoustic correlates of stress in Central Catalan and Castilian Spanish." *Language and Speech*

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*, 258-278.

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. *Journal of the Acoustical Society of America*, *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*, *12*, 123-144.

Payne, E., Post, B., Astruc, L., Prieto, P., Vanrell, M.M. (under revision). Measuring Rhythm in Child-Mother Interaction. *Language and Speech*.

Payne, E., Post, B., Astruc, L., Prieto, P., Vanrell, M.M. (2010). Rhythmic Modification in Child Directed Speech, in M. Russo (Ed.) *Atti del convegno di Prosodia, Gli universali prosodici: confronto e ricerche sulla modellizzazione ritmica e suller tipologie ritmiche*, Roma: Aracne Biblioteca di Linguistics,147-183.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Prieto, P. (in press). The Intonational Phonology of Catalan. In S.A. Jun (Ed.), *Prosodic Typology 2*. Oxford University Press: Oxford.

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

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002* (pp. 115-120). Aix-en-Provence.

Ramus, F., Nespor, 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 M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 337-342). Barcelona.

Recasens, D. & Espinosa, A. (2006). A dispersion and variability in Catalan vowels. *Speech Communication, 48*, 645-666.

Shen, Y., & Peterson, G. G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers, 9,* 1-36.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies*. London: Edward Arnold.

Russo, M., & Barry, W. J. (2008). Isochrony reconsidered. Objectifying relations between Rhythm Measures and Speech Tempo. *Proceedings Fourth Conference on Speech Prosody*, May 6-9 2008, Campinas, Brazil, pp. 419-422.

Toro, J. M., Trobalón, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, *6*, 131-136.

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics, 25*, 25-41.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, *27*, 171-206.

Vizcaino Ortega, F., Estebas-Vilaplana, E., Cabrera Abreu, M., Astruc, L. (2008). The phonological representation of edge tones in Spanish alternative questions. In A. Pamies, M. C. Amoros, & J. M. Pazos (Eds.), *Experimental Prosody. Special issue of Language Design*, *2*, 31-38. Granada: Granada Lingüística.

Wenk, B., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10*, 193-216.

White, L., & Mattys, S. L. (2007a). Calibrating Rhythm: First Language and Second Language Studies. *Journal of Phonetics*, *35,* 501-522.

White, L., & Mattys, S. L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró , & M. J. Solé (Eds.), *Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series)* (pp. 237–257). Amsterdam, Philadelphia: John Benjamins.

White, L., Payne, L., & Mattys, S. L. (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigário, S. Frota & M. J. Freitas (Eds.), *Phonetics and Phonology*. Amsterdam, Philadelphia: John Benjamins.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707-1717.

Yoon, T. J., Cole, J., & Hasegawa-Johnson, M. (2007). On the edge: acoustic cues to layered prosodic domains. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1264-1267). Pirrot GmbH: Dudweiler.

Appendix – Materials used in the reading experiment

Type 1. Predominantly CV-type utterances

The number in parenthesis at the end represents the number of syllables.

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 is of 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)

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 Ámsterdam 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 párquing 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).

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)

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)

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. Llueve 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)

Corresponding author:

Pilar Prieto
ICREA-Universitat Pompeu Fabra
Departament de Traducció i Ciències del Llenguatge
Campus de la Comunicació-Poblenou
Universitat Pompeu Fabra
Carrer Roc Boronat, 138
Despatx 53.710
08018 BARCELONA
Tel. 93 542 22 48
Fax. 93 542 16 17
E-mail: pilar.prieto@upf.edu