Article

Language and Speech

Language *and* Speech
1–27

DOI: 10.1177/0023830911417687
las.sagepub.com
SAGE

# Measuring Child Rhythm

**Elinor Payne**
University of Oxford, UK

**Brechtje Post**
University of Cambridge, UK

**Lluïsa Astruc**
The Open University, and University of Cambridge, UK

**Pilar Prieto**
Institució Catalana de Recerca i Estudis Avançats, and Universitat Pompeu Fabra, Spain

**Maria del Mar Vanrell**
Universitat Pompeu Fabra, and Universitat Autònoma de Barcelona, Spain

**Abstract**
Interval-based rhythm metrics were applied to the speech of English, Catalan and Spanish 2, 4 and 6 year-olds, and compared with the (adult-directed) speech of their mothers. Results reveal that child speech does not fall into a well-defined rhythmic class: for all three languages, it is more 'vocalic' (higher %V) than adult speech and has a tendency towards lower variability (when normalized for speech rate) in vocalic interval duration. Consonantal interval variability, however, is *higher* in child speech, particularly for younger children. Nevertheless, despite the identification of common, cross-linguistic patterns in child speech, the emergence of language-specific rhythmic indices is also clearly observable, even in the speech of 2 year-olds.

**Keywords**
Catalan, development, English, rhythm, Spanish

## 1 Introduction

This paper examines the extent to which 'rhythm metrics', based on extracting certain (predominantly durational) properties from the acoustic signal (cf. Ramus, Nespor, & Mehler, 1999; Low,

**Corresponding author:**
Elinor Payne, Phonetics Laboratory, University of Oxford, 41 Wellington Square, Oxford, OX1 2JF, UK
Email: elinor.payne@phon.ox.ac.uk

Grabe, & Nolan, 2000; Grabe & Low, 2002; Dellwo, 2006; White & Mattys, 2007), can usefully be applied to the developing phonology and phonetics of a child. While the efficacy of these metrics in capturing rhythmic differences in adult speech has been widely discussed and evaluated (see in particular a comparison of all metrics by White & Mattys, 2007), little attention has been paid to the *acquisition* of speech rhythm and the usefulness of metrics in investigating this. In this study, we apply a range of metrics based on acoustically defined intervals[1] to the speech of 2, 4 and 6 year-old Catalan, Spanish and English children to investigate how cross-linguistic differences detected in adult speech are acquired. In particular we address the following questions: How useful are the different rhythm metrics for a) testing cross-linguistic differences in child speech, and b) discriminating between child and adult speech? Are cross-linguistic differences in rhythmic indices already present in very early (age 2) multi-word utterances? What is the rhythmic 'profile' of child speech as depicted by the rhythm metrics and how does it differ from adult speech? Are these differences still observable in the speech of 6 year-olds?

### 1.1 Background on speech rhythm measurement

Attempts to define the acoustic bases of perceived differences in adult speech rhythm have a long and well-documented history (see White & Mattys, 2007, for an overview). An early thesis proposed that perceived differences in rhythm (i.e., between the perceived 'Morse Code' rhythm of for example Germanic languages and the perceived 'machine gun' rhythm of for example Romance languages) were dependent on these languages adopting (categorically) different timing strategies, with isochronous units coinciding with either syllables (hence the term 'syllable-timed') or stress-delimited feet (hence the term 'stress-timed') (Pike, 1945; Abercrombie, 1967). Many instrumental studies subsequently sought and failed to find any evidence of isochrony (cf. Lea, 1974; Pointon, 1980; Roach, 1982; Dauer, 1983). Nevertheless, though the acoustic basis is elusive, the percept of a distinction between two broad categories is strong, and empirically supported (Ramus, Dupoux, & Mehler, 2003, for example, show adult ability to distinguish *between* rhythmic categories, but not *within* them) and fuels on-going investigation.

An alternative explanation for the source of perceived rhythmic difference (see Bertinetto, 1981; Dauer, 1983, 1987; Roach, 1982; and Dasher & Bolinger, 1982) proposes that it emerges from distinct sets of phonological and phonetic properties found across languages, most notably the complexity of syllable structure and the presence versus absence of vowel reduction. In brief, so-called 'stress-timed' languages, like English, have a greater range of syllable structures, permitting complex codas and onsets, with heavier syllables more likely to attract stress, and also tend to reduce unstressed vowels (both durationally and quantitatively, see Delattre, 1966) when compared with vowels in stressed syllables. In contrast, in so-called 'syllable-timed' languages like Spanish, open syllables are far more common (many syllables are CV), and vowel reduction is much less evident. This hypothesis of rhythmic difference (which shall be referred to hereafter as the 'phonology-derived' hypothesis) in effect captures differences in the degree to which languages use duration to execute prosodic prominence. On this interpretation, the execution of prosodic prominence is not merely dependent on transparent timing strategies (hence the failure to find the basis of rhythm in isochrony) but interacts in a complex manner with the segmental string. To the extent that the proposed rhythm metrics 'work' (i.e., discriminate between perceived differences), it is because they reflect the acoustic *effects* of this interaction (i.e., variations in duration of vocalic and consonantal intervals), without having to take into account or model the component parts of this interaction (i.e., the language-specific structural and phonetic properties that result in these variations).

These claims find strong support in psycholinguistic studies which show that rhythmic differences are detectable from birth. Critical evidence (Nazzi, Bertoncini, & Mehler, 1998; see also Nazzi, Jusczyk, & Johnson, 2000; Nazzi & Ramus, 2003; Ramus et al., 2003) that infants attend to rhythmic differences from birth would appear to indicate that these differences reflect something that can be perceived before lexical or phonological analysis is available to the infant, and therefore can be captured objectively from the acoustic stream. From this, Ramus et al. (1999) argue that "the infant primarily perceives speech as a succession of vowels of variable durations and intensities, alternating with periods of unanalysed noise (i.e., consonants)" (p. 270). This insight led Ramus et al. (1999) to develop three measures of utterance rhythm which can be extracted purely on acoustic grounds, on the basis of vocalic and consonantal intervals: i) the standard deviation of vocalic intervals ($\Delta$V); ii) the standard deviation of intervocalic (i.e., consonantal) intervals ($\Delta$C); and iii) the proportion of total utterance duration which comprises vocalic intervals (%V). Application of these metrics to eight languages of perceived different rhythmic categories revealed a combination of $\Delta$C and %V or $\Delta$V to be the most useful in distinguishing categories (Ramus et al., 1999). The metric $\Delta$V also proved useful in discriminating between canonical 'stress-timed' and 'syllable-timed' languages, but not helpful in placing Polish (of supposedly 'intermediate' rhythm[2] class), since it was even lower for Polish than for 'syllable-timed' languages.[3]

Evidence that $\Delta$C and $\Delta$V are inversely proportional to between-language differences in speech rate[4] (e.g., Barry, Andreeva, Russo, Dimitrova, & Kostadinova, 2003; Dellwo & Wagner, 2003; Lee & Todd, 2004) led Dellwo (2006) to propose the rate-normalized metric VarcoC, and Ferragne and Pellegrino (2004) and White and Mattys (2007) to propose the rate-normalized metric VarcoV. VarcoC and VarcoV are variation coefficients calculated as the standard deviation of the interval duration in question ($\Delta$C or $\Delta$V) divided by the mean interval duration (mean C or mean V), and multiplied by 100.

A different approach to measuring the basis of rhythmic difference, but one nevertheless derived from acoustic intervals, is that proposed by Low et al. (2000), namely the application of the pairwise variability index (PVI). Rather than taking global 'vocalic-ness' and variability in that 'vocalic-ness', the PVI attempts to capture *sequential* differences in vocalic interval duration, and specifically between stressed and unstressed syllables. The motivation for looking at the sequential nature of the contrast is that prosodic prominence depends on syntagmatic contrast: what counts is a comparison with what has gone immediately before and with what lies immediately ahead. The PVI is calculated as the mean of differences between successive intervals, and is normalized (nPVI) for variability of speech rate by dividing by the sum of intervals. Grabe and Low (2002) also propose an intervocalic PVI (PVI-C), but warn against normalizing this. The rationale for this is that since the size and variability of intervocalic intervals largely reflect a language's phonotactics and since these are claimed to be an underlying source of that language's rhythmic properties, normalizing eliminates rhythmic difference.

With the proliferation of different metrics, one question that arises is how well they perform comparatively. There are in effect three main comparisons to consider here, namely i) that between the different category of metric (%V vs. the global interval variability measures vs. the sequential interval variability measures); ii) speech rate normalized versus non-rate normalized; and iii) vocalic versus consonantal measures. White and Mattys (2007) carried out a direct comparison of the usefulness of rhythm metrics in capturing perceived differences between 'stress-timed' English and Dutch on the one hand and 'syllable-timed' French and Spanish on the other (for adult speech). They found %V[5] and the rate-normalized vowel metrics (VarcoV and nPVI-V) to be the most effective overall (with a slight favoring of VarcoV over nPVI-V, also for within-category discrimination). They failed to find a significant correlation between %V and speech rate (see also Dellwo

& Wagner, 2003), and it would appear that this metric is fairly robust across different rates and there is thus no need for rate normalization. For interval variability metrics, however, White and Mattys (2007) found that while speech rate normalization increased the effectiveness of *vocalic* measures (VarcoV was found to have greater discriminatory power than $\Delta$V), this was not mirrored in the case of consonantal metrics, where speech rate normalization failed to facilitate discrimination between rhythmic types. Specifically, no evidence was found for systematic patterning between rhythmic classes when using the rate-normalized metric VarcoC. This is compatible with Grabe and Low's advocacy of leaving their intervocalic PVI measure (for consonants) raw (rPVI-C). In other words, normalizing speech variability over an utterance is deemed to be helpful for discriminating rhythm types when using vocalic interval variability measures, but unhelpful when using consonantal interval variability measures.[6]

Comparing consonant and vowel measures, in White and Mattys's study the former proved to be less effective or consistent in discriminating between rhythmic types, though non-normalized metrics fared better in this respect than normalized metrics, as already discussed. As White and Mattys (2007, p. 520) point out, the apparent comparative weakness of consonantal metrics may be due to the idiosyncrasies of the materials used (with variation in segmental content), and they cite previous studies (e.g., Ramus et al., 1999) that did achieve greater discriminatory success with consonant metrics.

Thus, in summary, between the different categories of metric it would appear that all three types (%V, global interval variability and PVI) discriminate fairly robustly between rhythm types, that speech rate normalization is recommended for vocalic variability measures but not for consonantal variability measures, and overall that vocalic variability measures would appear to be more reliable than consonantal variability measures.

## *1.2 Rhythm acquisition*

To date, research on sentence-level rhythm in acquisition has principally, though not exclusively, focused on perception. Indeed, as has been already noted, the development of interval-based metrics was to a large extent motivated and validated by the discovery that infants attend to rhythmic differences from birth (Nazzi et al., 1998). In a wider framework, it has been shown that infants' knowledge of native language prosody is detectable even by the last trimester before birth (Querleu, Renard, Versyp, Paris-Delrue, & Crepin, 1988; Hepper, Scott, & Shahidullah, 1993), and certainly precedes that of ambient segmental patterns (e.g., Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Luce, & Charles-Luce, 1994). There is evidence not only of infants' ability to *detect* the rhythmic pattern of their native language, but also of their *adapting* their own perceptual abilities to the relevant phonetic categories from an early age. For example, Skoruppa, Pons, Christophe, Bosch, Dupoux, Sebastián-Galles, Alves Limissuri, and Peperkamp (2009) report that by the age of 9 months, Spanish and French infants show language-specific tuning to abstract stress perception, with Spanish infants (whose ambient language has contrastive lexical stress) successfully keeping track of stress patterns when listening to pseudo-words, and French infants (whose ambient language has fixed stress) ignoring this dimension. A study by Höhle, Bijeljac-Babic, Herold, Weissenborn, and Nazzi (2009) shows the emergence of a language-dependent trochaic bias by 6 months (apparent in German, but not in French), strongly suggesting that rhythmic structure of the ambient language has a very early impact on infants' processing of stress information. This study confirms the results of an earlier study by Friederici, Friedrich, and Christophe (2007) which investigated the processing of rhythmic structure from event-related brain potentials and revealed that experience with German and French differentially affects the brain responses already in 4-month-old infants.

However, infants' abilities to detect certain regularities of input and process them in a particular way is one matter, their ability to integrate this perceptual knowledge with their early productions and reproduce these regularities is quite another. Infants have been shown to have the capacity for learning arbitrary distributional patterns (e.g., Saffran, Newport, & Aslin, 1996), and several studies show the beginnings of this integration by the end of the first year, with the emergence of language-specific properties in child productions (cf. Boysson-Bardies et al., 1986; Boysson-Bardies & Vihman, 1991). More recent studies suggest that infants' imitative abilities are apparent much earlier (cf. Kuhl & Meltzoff, 1996; Vouloumanos & Werker, 2007; Jones, 2009) and even appear to be active in shaping the melody of cries of newborn infants (Mampe, Friederici, Christophe, & Wermke, 2009).

That children use what they hear to inform what they go on to produce is arguably self-evident; how and when they do so is not, and thus we should not assume *a priori* that perceptual sensitivity to rhythm translates *immediately* or *directly* into production. In a study on segmental duration in babbling and first words – a property which is arguably a key building block en route to the acquisition of sentence-level rhythm – Vihman, Nakai, and DePaolis (2006) find evidence for an early mismatch between the evidence for finely tuned perceptual learning and what children actually produce, with adult-like rhythmic patterns apparently being acquired on a word-by-word basis. The authors conclude that "[a] more abstract knowledge of the rhythmic patterning appropriate to the native language can be expected to emerge only gradually from the combined perceptual and production knowledge or representation of increasing numbers of individual words, with further reorganisation needed once larger and more varied units begin to be combined and integrated – the challenge of prosodic learning at the syntactic level" (2006, p. 364). This would give us reason to expect that the rhythmic characteristics of early multi-word utterances do not constitute a seamless progression from the ability to produce isolated words with the correct (adult target) rhythmic pattern.

A further potentially confounding factor is that while (according to the 'phonology-derived' hypothesis) language-specific rhythm emerges from distinct combinations of phonological and phonetic properties, in child speech these properties are notoriously unstable and in flux (Vihman, 1996). The degree to which this flux deviates from a common path among children exposed to the same ambient language will affect how successfully the metrics reveal any systematicity in the emerging rhythmic profiles. For example, the gradual acquisition of the ability to produce complex consonant clusters may differ considerably between children, and even for a single child. Furthermore, the path a single child follows may not be *linear*, in the sense that any advances may theoretically be subject to temporary reversal. For example, with respect to final lengthening, Snow (1994) found that children start to control this after the onset of the multi-word stage (1;5–2;0), but then may experience a regression a few months later (see also Snow, 2006). These factors potentially militate against the metrics detecting any systematic patterns for child speech.

At the production level, early studies by Allen and Hawkins (1978, 1980) describe (particularly English) child speech as being more 'syllable-timed', due to a combination of factors related to the apparent primacy, in both perception and production, of stressed syllables. They cite earlier research showing that children show selective attention to stressed syllables (Blasdell & Jensen, 1970; Risley & Reynolds, 1970) and note a predominance of stressed syllables in early speech, which is "composed largely of reduplicated or partially reduplicated forms that are themselves short sequences of phonologically similar, unreduced monosyllables" (Allen & Hawkins, 1980, p. 233). This affects the overall rhythm of the speech: "2-year olds tend to use far fewer reduced syllables than do adults, so that their speech rhythm has fewer syllables per foot, or more beats per utterance; in short, it sounds more syllable-timed" (Allen & Hawkins, 1980, p. 231). The ability to reduce weak syllables in English is acquired relatively late and marks "an important step in the child's development toward adult phonological rhythm. […] [B]y the age of 4 or 5, rhythm becomes more adult-like, with increased rate and greater numbers of reduced nuclei" (Allen & Hawkins, 1978, p. 174).

Development towards adult rhythm also depends on phonetic mastery of other major correlates of stress: pitch, intensity and duration. More recent studies show that by 2 years, children have begun to exert some control over these in their productions of familiar words (Klein, 1984; Kehoe, Stoel-Gammon, & Buder, 1995). Nevertheless, there is evidence that f0 control is acquired earlier than timing or intensity (for a review see DePaolis, Vihman, & Kunnari, 2008). Astruc, Prieto, Payne, Post, and Vanrell (2009) analyzed one-word utterances controlled for stress produced by 2, 4 and 6 year-old English, Spanish and Catalan-speaking children[7] as part of a naming game. They found that while children as young as 2 years can control relevant intonation parameters such as pitch timing and height, they still do not control syllabic duration and still excessively lengthen final syllables. These findings support Allen and Hawkins' claim that adult target rhythm has not been acquired by age 2.

Later studies have also investigated cross-linguistic differences in either isolated word or sentence level rhythm (cf. Vihman, DePaolis, & Davis, 1998; Grabe, Watson, & Post, 1999; Vihman & Velleman, 2000; Vihman, Nakai, & DePaolis, 2006). Looking at the duration of disyllabic VCV sequences for two stages of development in American English, Welsh and French, Vihman et al. (2006) found no clear evidence of between-language differences at the four-word point.[8,9] They also found that all language groups made progress towards the adult model over time, but this was more evident in French and Welsh. This echoes findings by Grabe et al. (1999). Applying the PVI to vocalic intervals in the speech of French and English 4-year-old monolinguals, they found evidence that younger children produce speech that is 'more even timed' (i.e., more 'syllable-timed') than their parents', in both languages, though the difference was especially evident for English children. They concluded, in support of Allen and Hawkins (1978, 1980) that this 'more even' timing evident in child speech is the less marked, default setting, and that the production of 'less even' or 'stress-timing' is acquired later. Several studies comparing monolinguals and bilinguals (Spanish, English and German) have confirmed the greater tendency towards more even timing in the speech of younger infants (Kehoe & Lleó, 2002; Whitworth, 2002; Bunta & Ingram, 2007). Intriguingly, Kehoe and Lleó (2002) also report *less* even timing in the Spanish of a bilingual German-Spanish child, suggesting interference from German. This finding is problematic for the strong 'phonology-derived' hypothesis discussed above, unless it is also the case that the child's Spanish syllable structure and phonotactics were also influenced by German, a possibility not explicitly explored in that study.

# 2 Objectives and hypotheses of the present study

## *2.1 Outline of general objectives*

One aim of this study is to explore the efficacy of the rhythm metrics applied to child speech, in detecting both cross-linguistic differences and discriminating between child and adult speech. A further aim is to provide a more complete picture of the acquisition of language-specific rhythmic profiles (and in particular when these begin to emerge) and of the rhythmic character of child speech (and whether this is still evident at the age of 6). It seeks to achieve these aims by extending the age dimension (to 2, 4 and 6 years) and by applying both sets of interval-based metrics (both the PVI and those developed by Ramus, in addition to %V). The study investigates three languages

(English, Spanish and Catalan) which differ in certain phonological and prosodic properties, and are said to differ according to rhythmic class.

Two of these fall uncontroversially into two perceptually distinct classes. English (Southern Standard British variety) is canonically classed as 'stress-timed' in the rhythm literature. It displays a wide variety of syllable structure types, a high frequency of complex syllable types, strong quantitative vowel reduction in unstressed position[10] (Delattre, 1966), stress attraction to heavier syllables, and substantial final lengthening (Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992; Turk & White, 1999). Spanish (European variety) is canonically classed as 'syllable-timed'. It has a predominantly CV syllable structure, a low frequency of complex syllables, and displays very weak vowel reduction (Delattre, 1966), a weak correlation between stress and syllable weight, and weaker final lengthening than in English (Hualde, 2005; Ortega-Llebaria & Prieto, 2007).

Catalan holds a more ambiguous position in the schema of rhythmic classification by dint of its mixed phonological properties. Though its predominant syllable structure is CV (typical of 'syllable-timed' languages), it has a greater incidence of more complex syllables than Spanish, weak vowel reduction (see Mascaró, 2002, for a thorough review of this process), a moderate correlation between stress and syllable weight, and much less final lengthening than in English (Astruc & Prieto, 2006).

The apparent mix of 'syllable-timed' and 'stress-timed' properties in Catalan has led to claims, in line with a strong interpretation of the 'phonology-derived' account of rhythm, that Catalan is rhythmically 'intermediate' (Nespor, 1990). However, there is contradictory evidence from empirical studies. While Grabe and Low (2002), using the PVI, report an intermediate rhythm index, with metric scores clearly differing from Spanish, Ramus et al. (1999, 2003) find Catalan clusters with Spanish and other 'syllable-timed' languages such as Italian and French. Gavaldà-Ferré (2007) finds different results with different metrics: %V and $\Delta C$ class Catalan as 'syllable-timed' while the PVI classes it as intermediate. She also shows that different degrees of vowel reduction found in different dialects of Catalan have no impact on any rhythm scores, providing evidence of the relative independence of vowel reduction and durational reduction. Her findings echo those of Ramus et al.'s perceptual study that "suggest that vowel reduction in Catalan is not enough to make it depart from syllable-timing" (2003, p. 341). Prieto, Vanrell, Astruc, Payne, and Post (2008, submitted) also find evidence for Catalan clustering with Spanish, except for %V and $\Delta V$ which class it as intermediate. Thus, although the weight of the evidence is behind Catalan clustering with 'syllable-timed' languages, discrepancies clearly exist. The motivation for the inclusion of Catalan in the present study was to ascertain whether these relatively subtle differences from Spanish induce different patterns in children acquiring rhythm.

To conduct a valid comparison between adult and child speech, it was necessary that the adult and child speech material be as similar as possible, in both content and elicitation. Since the method of having subjects read prepared sentences, adopted by most previous studies, was not appropriate for young children, the material was elicited in a semi-spontaneous way for both groups, with adults and children responding to visual prompts as part of a structured game (described in more detail in the Method section below). Because differences in elicitation technique could theoretically introduce rhythmic variation, it was deemed desirable to confirm previously reported cross-linguistic rhythmic differences for the semi-spontaneous adult speech. In other words, we expected to find lower %V and higher variability scores (VarcoV, nPVI-V, $\Delta C$ and rPVI-C) for English when compared with Catalan and Spanish. Given the uncertainty surrounding a possible rhythmic difference between Catalan and Spanish, we expected to find much weaker metrical discrimination between these two languages. If claims of Catalan's intermediate profile were to be upheld, we would expect any observed differences to be in the direction of lower %V and higher variability scores than for Spanish.

## 2.2 Specific hypotheses

Our first hypothesis concerns the detection of cross-linguistic differences in child speech. If the metrics are able to detect cross-linguistic differences in adult speech rhythm, they should also allow us to detect cross-linguistic differences, if these exist, at the developmental stage. Grabe et al.'s (1999) study suggests they do exist (at least for French and English, and at the age of 4) and can be detected. However, that study only investigated nPVI-V. We extend that study by investigating also younger children (age 2 years) and older children (age 6 years), and by applying a full set of metrics (both vowel and consonant interval metrics, and both PVI and *Δ*/Varco). English remains the test language for 'stress-timing', but Spanish and Catalan are used instead of French for 'syllable-timing'. Our hypothesis is that cross-linguistic differences will be apparent in child speech across the metrics, that is English will show lower %V and higher interval variability, but that these differences will not be evident, or at most only weakly evident, for the very early speech (2 years). This prediction is based on findings reported above (Allen & Hawkins, 1978, 1980; DePaolis et al., 2008; Astruc et al., 2009) that children have not yet acquired vowel reduction or mastered the durational correlates of stress by the age of 2, and show a universal tendency to simplify syllables and reduce consonant clusters at an early age. A further part of this hypothesis is that the greatest difference will be between English and the other two languages (as for adult speech).

Our second hypothesis addresses the question of rhythmic distinction between child and adult speech. Child speech, particularly for younger children, may differ phonologically and phonetically from the adult target quite substantially, and our hypothesis is that this will be reflected in a distinction in metric scores, irrespective of the rhythmic category of the language in question, with the rhythmic profile of child speech proving more 'even-timed', or 'syllable-timed' than that of adult speech.

More specifically, in terms of relative 'vocalic-ness', we hypothesize that %V scores will be higher in child speech for any given language. This hypothesis is motivated principally by the later acquisition of syllable codas – young children typically delete consonants and/or simplify consonant clusters – resulting in a higher proportion of CV syllables. We hypothesize further that this distinction will be particularly evident for English (i.e., the difference between adult and child %V scores the greatest for English), given the findings of previous studies (Grabe et al., 1999; Vihman et al., 2006), and since there is a greater discrepancy in phonotactics between early productions and adult target for English than for the other languages examined.

In terms of vocalic interval variability, we hypothesize that child speech will show less variability than adult speech (lower scores for VarcoV and nPVI-V). This hypothesis is motivated by a number of observations, and has already been supported for nPVI-V for 4 year-olds (Grabe et al., 1999). Firstly, child speech has a smaller variety of syllable structure types, and therefore there will presumably be less variation in the duration of the vowel, since the number of consonants in the onset and especially the coda typically affect its duration. Secondly, as Allen and Hawkins have described (1978, 1980, and see the discussion in section 1.2), the ability to reduce weak syllables is acquired relatively late, and early speech is characterized by a predominance of stressed syllables. This means there should be less variability in nucleus-vowel duration as a result of prosodic prominence marking. Less variability in vocalic intervals would be compatible with a more 'syllable-timed' profile. As children begin to acquire the phonotactic and prosodic structure at the root of vocalic variability, their scores for VarcoV and nPVI-V are expected to increase. However, there is also a possibility that, in the transition towards adult characteristics, children go through a period of overshooting or undershooting articulations and hence produce even greater variability than adults.

In terms of consonant variability, a dimension that has not been specifically addressed in previous studies of child speech rhythm, our hypothesis is that child speech will be characterized by less variability in consonant intervals (lower scores for $\Delta$C and rPVI-C). The reasoning behind this is as follows. Although, as discussed, metrics based on variability of consonant interval duration are not as robust as vocalic metrics in discriminating between rhythmic classes for adult speech, various studies report higher variability in 'stress-timed' languages and this is claimed to be due to a wider variety of syllable structures.[11] If consonant intervals can typically often span 2 or more consonants, through the presence of simple and complex codas and onsets, greater variability in the duration of those intervals will result. Since children acquire codas and complex clusters later, we hypothesize that there will be less variability in child speech, especially at a younger age. More even timing in consonant intervals would also be compatible, at least in terms of adult rhythm classification, with higher %V and more even timing in vocalic intervals.

A second part of this hypothesis concerns the transition from the profile for child speech rhythm to the profile for adult speech rhythm. The former is not static, since over time children progressively acquire the whole set of phonetic and structural properties characteristic of their ambient language. While it is beyond the scope of the present paper to investigate the details and trajectory of this progression, either metrically or in terms of the speech material that underlies the metrics, our hypothesis is that by age 6, a sufficient number of language-specific phonetic and phonological parameters have been acquired to blur some of the edges in the distinction between adult and child speech. This would be in keeping with findings by Bunta and Ingram (2007) that show an early bias for even timing in bilingual Spanish and English-speaking children, and that this diminishes with time. As such, our hypothesis is that the characteristic rhythm profile for child speech (as mapped out for the first part of Hypothesis 2) will be less in evidence by age 6, and instead begin to show distinctly adult characteristics. In particular, it is hypothesized that by this age child speech in English will have begun to display recognizably 'stress-timed' characteristics and thus not be robustly distinguishable from adult speech in terms of the rhythm metrics.

### *2.3 Summary of hypotheses*

We hypothesized that:

1. H1 Child speech would show cross-linguistic divergence in rhythmic scores, except for age 2 years, when cross-linguistic metrical divergence would at best be very weak, principally because the phonotactic and prosodic parameters underpinning greater 'stress-timing' in English have not yet been fully acquired.
2. H2 Child speech rhythm scores would cluster around the 'syllable-timing' end of the continuum (more vocalic[12] and more even timing than equivalent adult speech), for all languages, and particularly for younger children. However, child-speech-specific characteristics would be weak by age 6, with English child speech displaying recognizably 'stress-timed' characteristics.

## 3 Method

### *3.1 Participants*

For each language, we recorded nine children interacting with their mothers (providing the child speech material), and made separate recordings of the mothers interacting with an adult interviewer (providing the adult speech material). The children fell into three age groups (i.e., there were three children in each age group for each language): 2, 4 and 6 year-olds. In total, 27 children and 22 adults were recorded.[13] Recordings were made respectively in the participants' homes in Cambridge, Madrid and Barcelona, 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.

### *3.2 Materials and elicitation*

The main goal was to elicit utterances that were, in terms of lexical targets, as similar as possible across adult and child speech. Our expectation was that the children, and in particular the younger children, would show some deletion of whole words and syllables, and would delete, modify or add individual segments. By seeking to prompt the same lexical targets, we hoped to elicit as segmentally comparable data as possible within the bounds of what can be expected for developmental speech. That way, most of the segmental differences would result from the developmental nature of the speech (intrinsically of interest) rather than randomly different material. Most previous studies on adult rhythm have not explicitly sought to control for segmental composition.

Twenty-three short utterances were elicited through the medium of a structured game, based on short, animated clips, shown on PowerPoint slides on a laptop screen (one utterance per slide). The animations showed simple, everyday scenes, which could easily be described in words that were highly familiar to the children. This was done to increase the likelihood of obtaining directly comparable speech samples across the age groups. For example, one scene showed a little girl blowing soap bubbles, another showed a little boy playing a trumpet. The mother was instructed to ask her child to describe what was happening in each clip, then praise the child for getting it right, and repeat what the child had said. A typical dialogue went thus:

Mother: *What's happening here? What's the little girl doing?*
Child: *[She's] blowing bubbles!*[14]
Mother: *That's right! She's blowing bubbles!*

The mothers were also recorded doing the same task, in the same role, interacting with an adult (the interviewer). Note that the adult speech used as a basis for comparison was thus *adult-directed* speech. This is important because it has been shown – and precisely for the languages under investigation here – that adults modify their speech rhythm when addressing children (Payne, Post, Astruc, Prieto, & Vanrell, 2010). While it could be argued that the child's 'immediate' target is actually the child-directed speech she interacts with, the current paper is concerned with how child speech rhythm differs from adult speech rhythm, and not with the extent to which it differs from its immediate target.

The total utterances extracted per age and language group are shown in Table 1. Only the target utterances were extracted for analysis. For example, for the answer '*she's blowing bubbles*', if the mother placed this in a broader context e.g., '*let me see … she's blowing bubbles*', only the target lexical items '*she's blowing bubbles*' were used in the analysis. This meant that child and adult utterances were as near as possible in length and in terms of segmental target, within the confines of the experimental design. The objective was not to elicit the exact same utterances for each subject (an impossible task cross-linguistically, and an unlikely outcome cross-age), but utterances of roughly comparable length (see examples in Table 3). Nevertheless, variation in utterance length occurred, as can be seen from Table 2, with English utterances on average shorter for all age groups than Catalan and Spanish utterances, and utterances by 2 year-olds on average shorter than utterances

**Table 1.** Total number of utterances per language for adult speech and each child age (mean per subject in parentheses).

| | Age 2 | Age 4 | Age 6 | Adult |
|---|---|---|---|---|
| English | 68 (17) | 86 (21.5) | 68 (17) | 226 (18.8) |
| Catalan | 62 (15.5) | 67 (16.8) | 69 (17.3) | 154 (12.8) |
| Spanish | 64 (16) | 89 (22.3) | 69 (17.3) | 138 (11.5) |

**Table 2.** Mean number of syllables per utterance per age and language group.

| | Age 2 | Age 4 | Age 6 | Adult |
|---|---|---|---|---|
| English | 3.11 | 4.79 | 4.40 | 4.49 |
| Catalan | 4.83 | 8.95 | 6.87 | 7.24 |
| Spanish | 4.19 | 6.69 | 8.45 | 8.65 |

**Table 3.** Example utterances cross-language and cross-age.

| Age | English | Spanish | Catalan |
|---|---|---|---|
| 2 | [pleɪjɪŋ mju:zɪk] | [koka muka] | [toka una muzika] |
| | *(playing music)* | *(toca música)* | *toca una música* |
| 4 | [peɪjɪn tjʌmhɪ] | [tokando una tɾompeta] | [tokant la tɾompeta] |
| | *(playing a trumpet)* | *(tocando una trompeta)* | *(tocant la trompeta)* |
| 6 | [ʃi:z pleɪjɪŋ ə tɹʌmpet] | [esta tokando la tɾompeta] | [esta tokant la tɾompeta] |
| | *(she's playing a trumpet)* | *(está tocando la trompeta)* | *(està tocant la trompeta)* |

by other age groups. An analysis of variance showed that the number of syllables did indeed vary with Language and Age group (there was a main effect of Language: $F(2, 37) = 30.34$, $p < 0.001$; and of Age: $F(3, 37) = 14.85$, $p < 0.001$), and that there was an Interaction of Language*Age ($F(6, 37) = 3.27$, $p < 0.05$). Since this could have potentially been a confounding variable, the data were checked for correlations between metric scores and utterance length. Although correlations between number of syllables and interval variability metrics were found[15] (Pearson's $r$ for VarcoV $r(1147) = –0.088$, $p < 0.01$; nPVI-V $r(1147) = –0.251$, $p < 0.001$; $\Delta$C $r(1160) = –0.193$, $p < 0.001$; rPVI-C $r(1160) = –0.258$, $p < 0.001$), the $r$ values were extremely small.

### *3.3 Analysis*

Vocalic intervals were segmented and labeled (start-points and end-points) from the waveform and spectrogram by a research assistant using *Praat* (Boersma & Weenink, 2007). The RA is a native speaker of English and studied Spanish language as part of her degree. The Catalan and Spanish labelings were revised by the three author native speakers. Vocalic and consonantal segmentation was performed with reference to standard criteria (see e.g., Peterson & Lehiste, 1960). The placement of a boundary between a vocalic interval and a consonantal interval was guided primarily by the presence of a sudden, significant drop in amplitude and a break in the formant structure, particularly F2. Marking the onset of consonants was facilitated by various cues, according to consonant manner. For instance, the onset of a fricative was marked at the start of high frequency energy (visible frication). The onset of a nasal consonant was marked by the presence of nasal formant structure and low amplitude in the waveform. Following White and Mattys (2007, p. 507), the placement of a boundary between a consonantal interval and a vocalic interval was guided primarily by the presence of a sudden, significant rise in amplitude and the onset of vocalic formant structure. Thus, any aspiration following stop release was included as part of the consonant interval. Following Ramus et al. (1999, p. 271), prevocalic glides were considered to be part of the consonantal intervals; postvocalic glides part of the vocalic intervals. Following Grabe and Low (2002) and White and Mattys (2007), utterance-final intervals were not excluded from analysis, despite likely lengthening effects in this position, since, as White and Mattys (2007, p. 507) point out, "the locus and extent of final lengthening and other prosodic lengthening processes […] may be language specific and may contribute to the overall perception of cross-linguistic differences in rhythmicity."

Although child productions of target words, particularly with younger children, may vary considerably from adult forms, with frequent segment changes and deletions, this did not render the task of vocalic and consonantal interval boundaries any more difficult. Interval boundaries were placed according to actual articulations, not targets, precisely to capture the differences in phonotactics and syllable structures of child speech predicted to influence rhythm. Thus, the segmentation of 'trumpet' as for example [tʃʌ̰pɪ] is no harder than that of [tɹʌ̰mpɪt], just different.

Vocalic and intervocalic (consonantal) intervals were then extracted using a Praat script, and the following metrics were calculated (after Grabe & Low, 2002, and White & Mattys, 2007):

i) Measure of 'vocalic-ness': %V
ii) Measures of variability in vocalic interval duration (rate normalized): VarcoV and nPVI-V
iii) Measures of variability in consonantal interval duration (not rate normalized): $\Delta$C and rPVI-C.

Rate-normalized metrics were chosen for vocalic intervals and non-normalized metrics were chosen for consonant intervals since, as has been discussed in section 1.1, these have been shown to have greater discriminatory power for adult speech. In addition, speech rate was calculated by dividing the number of vocalic intervals (roughly equivalent to the number of syllables) by the total time in seconds of each utterance.

# 4 Results and discussion

We carried out a series of univariate by-subject (F1) and by-item (F2) analyses of variance on all of the metric scores and speech rate to test the following: i) confirmation of cross-linguistic differences in semi-spontaneous adult speech (to establish a baseline with which to compare child speech) (see section 4.1); ii) cross-linguistic differences in child speech (Hypothesis 1) (see section 4.2); iii) differences between child and adult speech (Hypothesis 2) (see section 4.3); iv) differences in speech rate (see section 4.4). We only report results that attained significance in both F1 and F2. We report Bonferroni-corrected post hoc comparisons based on the (more stringent) subjects analyses: as is customary, the number of paired comparisons always equaled the total number of comparisons possible between the levels of the variable in question.

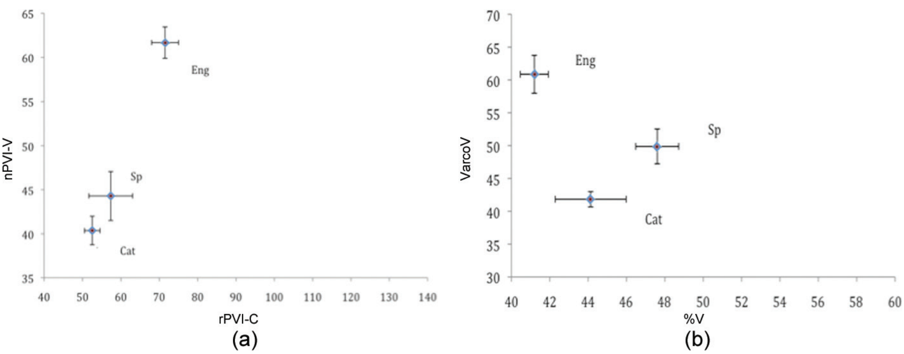
### *4.1 Confirmation of cross-linguistic differences in semi-spontaneous adult speech*

Since unlike most previous studies on speech rhythm the material in the present study was semi-spontaneous, in order to conduct a valid comparison between the child and adult speech data, it was first necessary to confirm whether previously reported cross-category differences could be replicated in the adult speech data.

For adult speech, in line with previous studies, there was found to be a main effect of Language for all metrics (%V: $F1(2, 19) = 12.96$, $p < 0.001$; $F2(2, 204) = 17.13$, $p < 0.001$; $\Delta$C: $F1(2, 19) = 10.13$, $p = 0.001$; $F2(2, 204) = 14.71$, $p < 0.001$; VarcoV: $F1(2, 19) = 9.73$, $p = 0.001$; $F2(2, 204) = 19.64$; $p < 0.001$; nPVI-V: $F1(2, 19) = 41.81$, $p < 0.001$; $F2(2, 204) = 48.05$, $p < 0.001$; rPVI-C: $F1(2, 19) = 5.30$, $p < 0.05$; $F2(2, 204) = 9.96$, $p < 0.01$).

Bonferroni-corrected post hoc comparisons were performed on Language (i.e., between all possible pairs of languages, which equaled 3 pairwise comparisons). These comparisons revealed that the only consistently robust difference is between English on the one hand and Spanish and Catalan on the other, while the latter two are more similar. For most metrics, English is distinct from Catalan ($\Delta$C: $p < 0.01$; rPVI-C: $p < 0.05$; VarcoV: $p < 0.001$; nPVI-V: $p < 0.001$), and from Spanish (%V: $p < 0.001$; $\Delta$C: $p < 0.05$; nPVI-V: $p < 0.001$). As we would expect for a 'stress-timed' language, English shows greater variability in both consonant and vocalic interval duration (see Figures 1a and 1b. Actual means are given in Table 4 in the Appendix), and a lower %V (see Figure 1b). The choice of parameters represented graphically in Figures 1a and 1b follows White and Mattys (2007) for their adult rhythm data, in plotting %V against VarcoV and rPVI-C against nPVI-V, and reproduces a very similar pattern to the results for Spanish and English in that study[16] (compare with their Figures 1 and 2, pp. 511–512). The results are also consistent with those for speech material obtained from the same set of English, Catalan and Spanish subjects as the present study, but reading carefully prepared sentences of equal length (see Prieto et al., submitted). This reassured us that the elicitation technique adopted had not introduced any spurious rhythmic effects.

The distinction between Catalan and Spanish is extremely weak. However, and in line with the findings of Prieto, Vanrell, Astruc, Payne, and Post (2008) and Prieto et al. (submitted) for read speech from controlled material, %V is significantly lower in Catalan than in Spanish ($p < 0.05$), as would be expected if Catalan were indeed less 'syllable-timed' than Spanish. Nevertheless, in all




**Figures 1a and 1b.** Distribution of Catalan, Spanish and English adult speech over the rPVI-C, nPVI-V plane (left) and over the %V, VarcoV plane (right). Error bars represent one standard error around the mean.

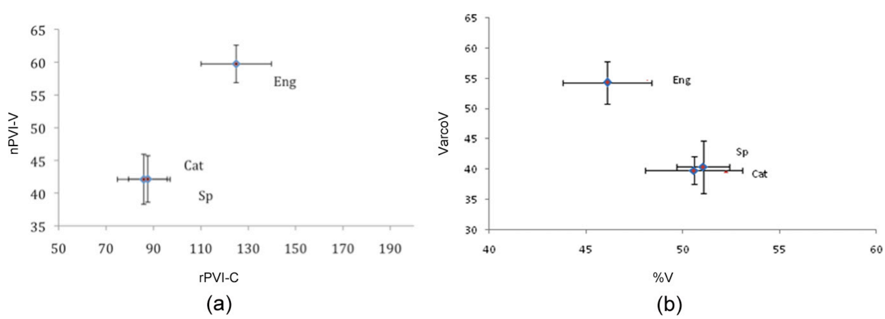
other respects, Catalan and Spanish clustered tightly together. This absence of a clear-cut difference between Catalan and Spanish is consistent with the contradictory results of previous studies, as discussed in section 2.

## 4.2 Hypothesis 1: Cross-linguistic rhythmic divergence in child speech

Our first hypothesis (H1) was that child speech would show cross-linguistic divergence in rhythmic scores except for age 2 years, when cross-linguistic metrical divergence would at best be very weak. Our results show that cross-linguistic differences are indeed detectable in child speech. Univariate ANOVAs were run on the child metric scores, with Language and Child Age as factors. As for adult speech, there was shown to be a main effect of Language on all metrics (%V: $F1(2, 18) = 4.13$, $p < 0.05$; $F2(2, 198) = 12.52$, $p < 0.01$; $\Delta$C: $F1(2, 18) = 11.18$, $p = 0.001$; $F2(2, 198) = 27.87$, $p < 0.001$; VarcoV: $F1(2, 18) = 13.81$, $p < 0.001$; $F2(2, 198) = 18.86$, $p < 0.001$; nPVI-V: $F1(2, 18) = 19.78$, $p < 0.001$; $F2(2, 198) = 31.60$, $p < 0.001$; rPVI-C: $F1(2, 18) = 9.33$, $p < 0.01$; $F2(2, 198) = 28.44$, $p < 0.001$).

Bonferroni-corrected post hoc comparisons (3 pairwise comparisons) showed that, as for adult speech, the only robust difference is between English on the one hand and Catalan and Spanish on the other. The difference between English and Spanish was significant for all metrics (%V: $p < 0.05$; $\Delta$C: $p < 0.01$; VarcoV: $p = 0.001$; nPVI-V: $p < 0.001$; rPVI-C: $p < 0.01$), and similarly the difference between English and Catalan was significant for all metrics (%V: $p < 0.05$; $\Delta$C: $p < 0.01$; VarcoV: $p = 0.001$; nPVI-V: $p < 0.001$; rPVI-C: $p < 0.01$). As with adult speech, English child speech shows greater variability in both consonant and vocalic interval duration, a lower %V than Spanish and Catalan, and a lower speech rate than the other two languages (see Figures 2a and 2b and Table 5 in the Appendix). The difference between Catalan and Spanish, already extremely weak in adult speech, is non-existent for child speech, with no significant differences found.

The second part of the hypothesis predicted that cross-linguistic differences would be weak or non-existent at the youngest age tested (age 2). Separate univariate ANOVAs on the metric scores for 2 year-olds, with Language as a factor showed, however, that the cross-linguistic differences are already robust for this age, at least for interval variability measures. There was a main effect of Language for all the metrics (VarcoV: $F1(2, 6) = 6.53$, $p < 0.05$; $F2(2, 66) = 5.72$, $p < 0.01$; nPVI-V: $F1(2, 6) = 6.34$, $p < 0.05$; $F2(2, 66) = 7.45$, $p = 0.001$; $\Delta$C: $F1(2, 6) = 11.69$, $p < 0.01$; $F2(2, 66) =$



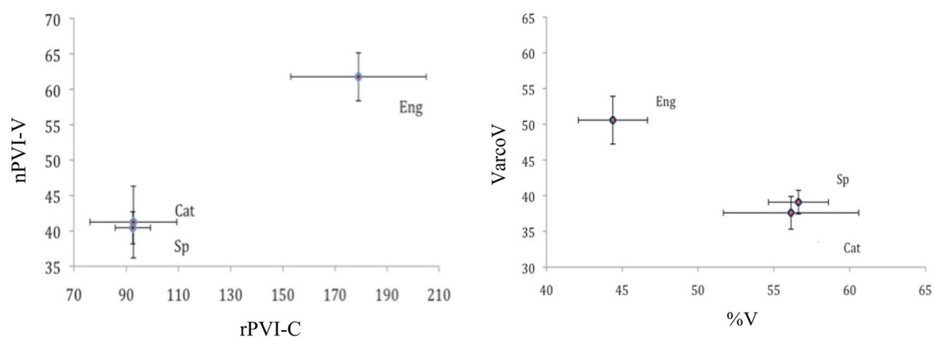

**Figures 2a and 2b.** Distribution of Catalan, Spanish and English child speech over the rPVI-C, nPVI-V plane (left) and over the %V, VarcoV plane (right). Error bars represent one standard error around the mean.

27.50, $p < 0.001$; rPVI-C: $F1(2, 6) = 7.65$, $p < 0.05$; $F2(2, 66) = 23.84$, $p < 0.001$), except for %V, where significance was attained only for F2, and was marginal in F1 ($F1(2, 6) = 4.42$, $p = 0.066$; $F2(2, 66) = 17.64$; $p < 0.001$).

Bonferroni-corrected post hoc comparisons (3 pairwise comparisons) revealed that Catalan and English were distinct for consonantal variability metrics ($\Delta$C: $p < 0.05$; rPVI-C: $p < 0.05$) and that Spanish and English were distinct for all variability metrics (VarcoV: $p < 0.05$; nPVI-V: $p < 0.05$; $\Delta$C: $p < 0.05$; rPVI-C: $p < 0.05$). In all cases, English variability scores were higher than Spanish and Catalan. Catalan and Spanish were not distinct for any of the metrics (see Figures 3a and 3b).

It would appear then that cross-linguistic differences in interval variability (though not, it would seem, in %V) observed in adult speech are already present in the speech of 2 year-olds, contrary to what we hypothesized. We cannot, however, assume that these differences do not fluctuate or change over the acquisition period. Indeed, the ANOVAs run on the scores for all child ages also revealed a main effect of Child Age for all metrics (though not always in both *F1* and *F2*): %V: $F1(2, 18) = 4.71$, $p < 0.05$; $F2(2, 198) = 14.27$, $p < 0.001$; VarcoV: $F1(2, 18) = 1.64$, ns; $F2(2, 198) = 3.30$, $p < 0.05$; nPVI-V: $F1(2, 18) = 4.12$, $p < 0.05$; $F2(2, 198) = 0.17$, ns; $\Delta$C: $F1(2, 18) = 3.39$, $p = 0.056$; $F2(2, 198) = 10.82$, $p < 0.001$; and rPVI-C: $F1(2, 18) = 7.08$, $p < 0.01$; $F2(2, 198) = 22.29$, $p < 0.001$. Furthermore, there was a significant interaction of Language*Child Age for %V and most of the consonantal variability metrics (%V: $F1(4, 18) = 3.07$, $p < 0.05$; $F2(4, 198) = 9.60$, $p < 0.001$; $\Delta$C: $F1(4, 18) = 2.56$, ns; $F2(4, 198) = 7.61$, $p < 0.001$; rPVI-C: $F1(4, 18) = 3.37$, $p < 0.05$; $F2(4, 198) = 11.93$, $p < 0.001$). This suggests that not only do relative vocalic-ness and variability in consonant intervals vary with child age, but also that they do so in different ways for different languages. By contrast, there was no interaction of Language*Child Age on vocalic interval variability metrics, suggesting cross-linguistic differences are more robustly present from an earlier age.

A brief examination of mean scores (reproduced in Table 6, Appendix) provides the following broad picture. Firstly, between the ages of 2 and 6, Catalan and Spanish become less 'vocalic'. Secondly, consonant variability follows a consistently downward trend over the age range for all three languages, but much more dramatically in English, particularly between ages 2 and 4. Mean scores in rPVI-C for English progress as follows: 179 > 102 > 94. Compare this with a much more limited reduction in Spanish (93 > 89 > 81) and Catalan (93 > 89 > 76). In other words, though they




(a) (b)

**Figures 3a and 3b.** Distribution of Catalan, Spanish and English child speech at age 2 years, over the rPVI-C, nPVI-V plane (left) and over the %V, VarcoV plane (right). Error bars represent one standard error around the mean.

may start out strong, cross-linguistic differences in consonant variability may not be consistently robust over this period.

To summarize the findings for Hypothesis 1: the first part of the hypothesis – that cross-linguistic differences are present in child speech – is broadly supported, while the second part – that these differences would be weak or non-existent at age 2 years – is not. Variability indices of adult speech rhythmic class distinctions are observed to be robustly present in child speech already by the age of 2 years, with Catalan and Spanish child speech exhibiting more ‘even-timed’ characteristics than English. Cross-linguistic differences in consonantal variability and %V appear to fluctuate to some extent over the age range, with %V diminishing in Catalan and Spanish and consonant variability decreasing at different rates. In contrast, vocalic variability appears to be fairly stable over this age range. This suggests that the cross-linguistic differences may not be consistently strong for all parameters. A detailed study of the phonological development of the children’s speech over this time period may help to chart and explain the trajectory of the metrics over this time period.

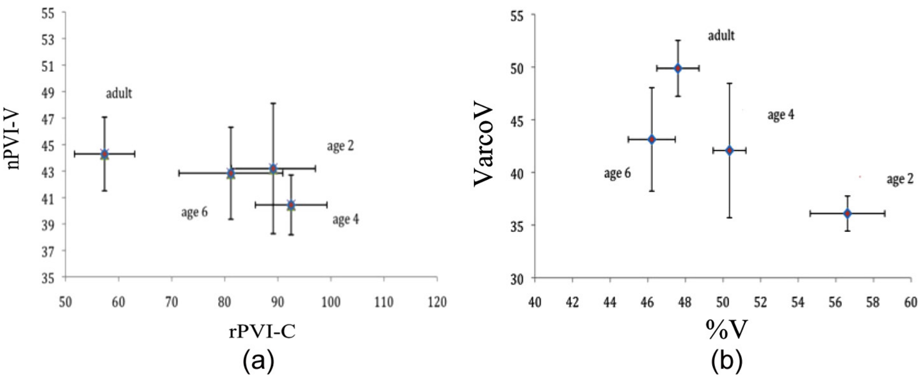
### *4.3 Hypothesis 2: Rhythmic properties of child speech*

Our second hypothesis (H2), relating to rhythmic type, was that child speech rhythm scores would be more characteristic of ‘syllable-timing’ (i.e., higher %V and lower interval variability) than adult speech, but that these child-speech-specific characteristics would be weak by age 6. We report our findings in three subsections: i) vocalic-ness (%V); ii) vocalic interval variability (VarcoV and nPVI-V); and iii) consonant interval variability ($\Delta$C and rPVI-C). A univariate ANOVA was run on the child and adult scores, with Age as a factor (4 levels of age, i.e., age 2, age 4, age 6 and adult) for each metric. Bonferroni-corrected post hoc comparisons were performed on AGE (i.e., between all possible pairs of ages, totally 6 pairwise comparisons), with only those between age 6 and adult reported.

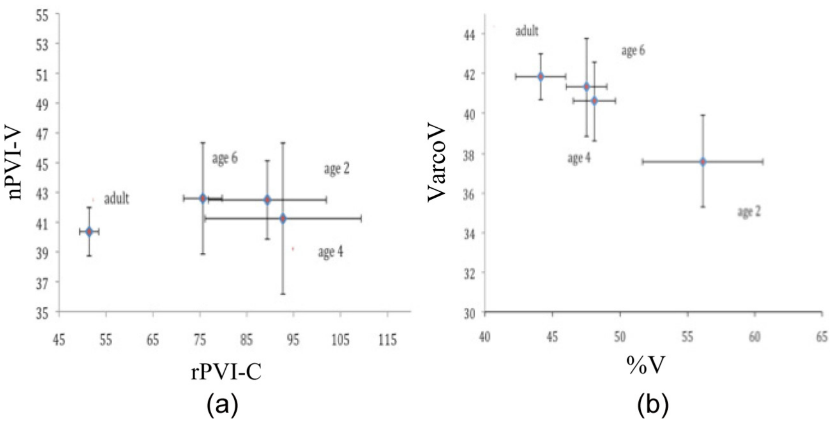
*4.3.1 Vocalic-ness (%V).* Age was found to have a main effect on %V, $F1(3, 37) = 14.40$, $p < 0.001$; $F2(3, 401) = 28.41$, $p < 0.001$. Bonferroni-corrected post hoc comparisons revealed that adult scores were significantly different from all groups of child scores except age 6 (adult versus age 2: $p < 0.001$; adult versus age 4: $p < 0.001$; adult versus age 6: ns). As predicted, %V scores were higher in child speech for these ages (2 and 4 years) than in adult speech in all languages (see %V scores in Figures 4b, 5b and 6b). These results are compatible with the hypothesis that child speech is more characteristically ‘syllable-timed’ in its rhythm than adult speech, but that by age 6 this contrast is much weakened. Indeed, it would appear at least in Spanish and English that the contrast is no longer present (see Figures 4b and 6b).

*4.3.2 Vocalic interval variability.* For normalized scores, Age was found to have a main effect only for VarcoV, $F1(3, 37) = 2.89$, $p < 0.05$; $F2(3, 401) = 8.04$; $p < 0.001$. However, Bonferroni-corrected post hoc comparisons revealed that child VarcoV scores were significantly lower than adult scores only at age 2 (adult versus age 2: $p < 0.05$; adult versus age 4: ns; adult versus age 6: ns) (see VarcoV scores in Figures 4b, 5b and 6b). Again, these results are compatible with the hypothesis that child speech is more characteristically ‘syllable-timed’ in its rhythm than adult speech, but that by age 6 this contrast is no longer present.

*4.3.3 Consonant interval variability.* Age was found to have a main effect on both metrics ($\Delta$C: $F1(3, 37) = 17.62$, $p < 0.001$; $F2(3, 402) = 53.15$, $p < 0.001$; rPVI-C: $F1(3, 37) = 24.54$, $p < 0.001$;



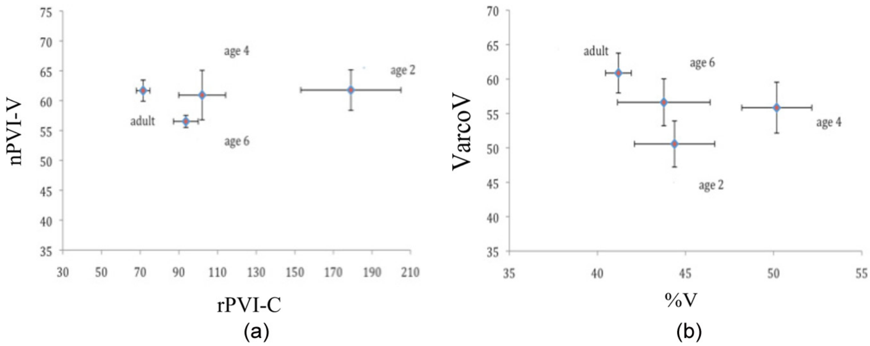

**Figures 4a and 4b.** Spanish data: Distribution of adult speech and child speech over the rPVI-C, nPVI-V plane (left) and over the %V, VarcoV plane (right). Error bars represent one standard error around the mean.




**Figures 5a and 5b.** Catalan data: Distribution of adult speech and child speech over the rPVI-C, nPVI-V plane (left) and over the %V, VarcoV plane (right). Error bars represent one standard error around the mean.

*F2*(3, 402) = 78.83, $p < 0.001$). Bonferroni-corrected post hoc comparisons revealed that child scores were significantly higher than adult scores for all child ages (for each metric, adult versus age 2: $p < 0.001$; adult versus age 4: $p = 0.001$; adult versus age 6: $p = 0.05$) (see also Figures 4a, 5a and 6a). This observed higher consonant interval variability does not support our hypothesis of greater syllable-timing in child speech. However, as we will suggest in section 5, this may be due to different factors underlying variability in vocalic and consonantal interval duration. With respect to the second part of the hypothesis, however, 6 year-olds are indeed closer to the adult target, as predicted. Nevertheless, as is evident from Figures 4a–6a, even at age 6 years, consonant variability is still considerably higher than for adults.

*4.3.4 Summary of child speech rhythmic characteristics*. The results provide evidence that, regardless of the language spoken, child speech is more vocalic and shows less variability (when normalized




**Figures 6a and 6b.** English data: Distribution of adult speech and child speech over the rPVI-C, nPVI-V plane (left) and over the %V, VarcoV plane (right). Error bars represent one standard error around the mean.

for speech rate) in vowel interval duration than the speech of adults directed to other adults. These are characteristics more commonly associated with so-called 'syllable-timed' rhythm, and are in keeping with the findings of previous studies for 4 year-olds (cf. Grabe, Watson, & Post, 1999). However, there is an apparent conflict between consonantal and vocalic scores: consonant intervals show significantly *higher* variability in child speech, more characteristic of 'stress-timed' rhythm. Closer examination of the speech of 6 year-olds compared with adult speech reveals, however, that by this age the differences in %V and vocalic variability are no longer significant in any of the languages investigated, and suggests that substantial progress has been made towards adult targets, in terms of acquiring both appropriate proportions of vocalic stretches and appropriate degrees of durational variability in those vocalic stretches. The same cannot be said for consonant interval variability, since though progress towards adult targets is evident, variability remains considerably higher in child speech than in adult speech even at this age.

### 4.4 Speech rate

The fact that in many of the variability results reported here there is a difference between normalized and non-normalized scores strongly suggests speech rate is an important factor for consideration. A univariate ANOVA was run on speech rate scores with Age (child versus adult) and Language as factors. There was shown to be a significant main effect of Age, $F1(1, 43) = 120.85$, $p < 0.001$; $F2(2, 408) = 784.50$, $p < 0.001$, with speech rate slower in child speech than in adult speech, across languages, as Figure 7 shows. There was also a significant main effect of Language, $F1(2, 43) = 12.08$, $p < 0.001$; $F2(2, 408) = 64.31$, $p < 0.001$. There was no interaction between Age and Language. Post hoc comparisons revealed a significant difference between English and Catalan ($p < 0.01$) and between English and Spanish ($p < 0.01$), with speech rate slower in English in both cases. There was no significant difference between Catalan and Spanish.

## 5 Summary and general discussion

### 5.1 The emergence of cross-linguistic rhythmic differences in child speech

A principal finding of this study is that cross-linguistic rhythm metric distinctions clearly detectable in adult speech are also clearly observable in child speech, even at the age of 2 years (in the case of interval variability). The differences between English and Catalan/Spanish are particularly robust: child speech in English has a slower speech rate, higher variability in consonant and vocalic interval duration and (at least beyond the age of 2 years) lower %V than child speech in Spanish and in Catalan. In other words, *according to the metrics*, English child speech is already less 'even-timed' (more 'stress-timed' rhythm) than Spanish and Catalan child speech. On the other hand, any putative distinction between Catalan and Spanish, the evidence for which is already very weak in adult speech, is not yet detectable in child speech for the ages investigated.

There is evidence that the *degree* of difference diminishes somewhat over the age range investigated, as Catalan and Spanish child speech becomes slightly less vocalic, and consonant interval variability is attenuated in all three languages but more rapidly in English child speech. Interestingly, since a decrease in %V automatically means an increase in %C, this means that the proportion of consonants in child speech *increases* (in Catalan and Spanish) as *variability* in consonant intervals *decreases*. At first sight this might appear surprising since, all other things being equal, an expected outcome of there being more consonants is *greater* variability in consonant duration. Indeed, according to the 'phonological approach' to rhythm, this is the rationale behind 'stress-timed' languages like English having greater variability in consonant intervals, since typologically such languages are observed to permit more complex syllables, that is syllables with codas (as shown also in their lower %V).

However, at the developmental stage, there is another factor to be taken into consideration, and that is that variability in consonant duration is also a product of immature phonetic abilities, and as such, as these abilities develop, variability will decrease, irrespective of whether the child is yet achieving syllable structure targets. Thus, phonetic and phonological abilities may be interacting in a complex manner with respect to their impact on the rhythm metrics. Hence, a low %C coupled with high variability in consonant duration may reflect a stage when children are wavering both in their attainment of the correct syllable structure and in the timing of the consonants they actually do produce. In other words, this is a stage when phonetic uncertainty parallels phonological uncertainty. As these abilities develop, %C increases and variability decreases, as seen in the speech of slightly older Spanish and Catalan children. The picture for English is slightly different. The fact that %V/%C do not vary as much over this period of development, but consonant interval variability decreases dramatically, would suggest that attainment in adult phonological structure makes critical advances at an earlier stage than the fine-tuning of phonetic ability, that is to say that children start producing the (more frequent) codas of English before they have fine-tuned their phonetic abilities. These broad patterns of development, and in particular the co-development of

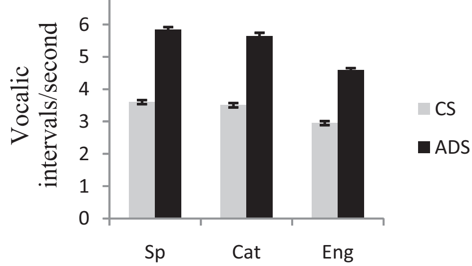



**Figure 7.** Mean speech rate in vocalic intervals/second in Spanish, Catalan and English child speech and adult speech. Error bars represent one standard error around the mean.

phonetic and phonological abilities, are important pointers for further investigation. The trajectory of metric scores from age 2 to age 6 in each language is undoubtedly a complex one, and is likely to be closely linked to language-specific patterns in the acquisition of segmental and syllabic abilities and the detail of prosodic lengthening. Further investigation into these properties and how they are manifested in the child speech material is needed in order to understand the developmental aspect of rhythm acquisition more fully.

### *5.2 The rhythmic characteristics of child speech*

This study has shown that child speech is characterized by a higher %V, lower vocalic interval variability and higher consonant variability than adult speech. This does not support the hypothesis that child speech *categorically* clusters around the more 'syllable-timed' end of the rhythm continuum, since higher consonant interval variability is reported to be more characteristic of 'stress-timing'. It does, however, suggest certain commonalities in the development of child speech, irrespective of language, and irrespective of the cross-linguistic distinctions clearly evident even at this early stage.

The data reported here show that early vowel productions, across languages, are characterized by more even timing, in contrast to early production of consonant intervals, which are characterized by *less* even timing than adult speech. There are a couple of apparent anomalies in these observed patterns. Firstly, why should vocalic intervals become *more* variable, but consonant intervals become *less* variable, with age? One might reasonably expect these properties to pattern in a similar way, even allowing for the possibility of one lagging behind the other. Secondly, one might also reasonably expect consonant interval duration to become *more* variable as the child produces more consonant codas and more consonant clusters.

One possible explanation that would fit the data is that different factors underlie vocalic and consonantal timing variability in developing speech. While basic vowel production from a phonetic perspective may be mastered early on, variability in vocalic *timing* would be seen to reflect language-specific prosodic patterns in the ambient language and is acquired later on. In other words, since it encodes prosodic structure, variability in vocalic duration is linguistically useful. On this interpretation, the main task for the child following the period observed in this investigation is to acquire and master this variability. In contrast, consonant production from a phonetic perspective is mastered later on, resulting in greater variability in timing in early productions. In other words, this variability is extraneous. For consonants, therefore, the main challenge for the child is to master control *over* this extraneous variability, and thus reduce it.

This interpretation would be in keeping with findings by Allen and Hawkins (1980, p. 236) who report different patterns for vowels and consonants in converging with adult norms. Specifically, although "vowels show mature patterns of relative and absolute duration quite early, by around 3 years for heavy syllable nuclei and by 4 or 5 years for light nuclei", they take much longer to master accent-related rhythmic patterns, and "their production of the full system of stress distinctions may remain inaccurate in some respects until the age of 12". In contrast, "consonant durations show a much longer developmental pattern, characterised by great variability within and between children and among consonant types and contexts. Target consonant duration and variability around that target do not stabilise for the average child until about age 10 or later." In other words, localized vocalic timing is mastered early on, but variability from longer domain prosodic structure occurs later, while control over consonant timing more generally occurs later on.

These patterns reveal cross-linguistic commonalities in the way children acquire mastery over vocalic and consonantal timing. However, the data indicate that children also show sensitivity to cross-linguistic phonotactic differences early on and reflect this in their own early productions: English child speech displays a high degree of consonant interval variability at age 2. As child age increases, these strongly language-specific characteristics appear to be attenuated, though do not of course disappear. The proportion of consonantal intervals in Catalan and Spanish is shown to increase (since there is a decrease in %V) as child age increases, and this is accompanied by a decrease in variability in consonant intervals. This would be evidence of a parallel phonological and phonetic development: as children start to attain adult target syllable structure, they also start to reduce the variability in their consonant production (at least in terms of duration). Thus, contrary to what a strong phonological approach to rhythm would presumably predict, the introduction of more consonants (a decrease in %V) in the speech of children does not automatically equate with greater variability in consonant interval duration. This is because, in child speech, consonant interval variability is not dependent *only* on phonotactic structure, but also on degree of mastery in consonant production.

Overall, the picture to emerge is one of the child facing a variety of different, possibly conflicting, challenges. On the one hand, there is the general challenge of mastering vowel and consonant production, and the production of these in sequences. This results, cross-linguistically, in certain commonalities when compared with adult speech: a higher proportion of vocalic intervals (most likely reflecting a greater incidence of simpler syllable structures); lower variability in the duration of those vocalic intervals (possibly due to the fact that structurally-dependent variation has not yet been acquired); and higher variability in the duration of consonant intervals (consonant production requiring more advanced articulatory skills). A likely explanation for these commonalities is that, for this early age range, children's productions are particularly influenced and guided by biomechanical constraints (Davis, MacNeilage, & Matyear, 2002). However, these can be expected to weaken over time, just as language-specific patterns strengthen (see Vihman, Kay, Boysson-Bardies, Durand, & Sundberg, 1994) and the child starts to tune her productions to the phonetic detail of her ambient language.

Attuning to adult targets can be seen to occur relatively early on, since the study has shown that by age 6, the rhythmic profile of child speech is well on the way to being isomorphic with that of adult speech, at least insofar as the vocalic components of this profile are concerned. This suggests that significant progress is made over this period in acquiring language-appropriate degrees of vocalic variability to approximate towards the relevant adult targets. Progress is also made in reducing consonantal variability, but at a slower pace, and, in line with Allen and Hawkins (1980), would appear to require a much longer developmental span. Since these targets differ, the precise nature of the progress achieved in reaching these targets also differs. The challenge to reduce consonant variability is greater for those acquiring Spanish or Catalan than for those acquiring English, since adult levels of consonant variability in the former are lower. Indeed, as Figures 4–6a indicate, by 6 years, children are closer to adult consonant targets in English than in either Spanish or Catalan. Conversely, for those acquiring English, in which adult levels of consonant variability are relatively high, the greater challenge is to acquire higher levels of vocalic variability.

In short, the rhythmic indices of early child productions *do* differ from those of adult speech, and in a way that is common across the languages investigated. However, these indices cannot be categorically classed as more 'syllable-timed'. Furthermore, though the cross-linguistic child productions display common properties, they also already display influence of the ambient language. It cannot, therefore, be said that children start out with a 'default' rhythm which then takes on language-specific properties, rather that rhythmic indices of both 'child speech' and the ambient language co-exist from an early age.

# 6 Conclusions and next steps

The interval-based metrics applied in this study have proved useful in that they yield what is perhaps a surprisingly[17] clear picture of cross-language and cross-age differences in the acoustic indices of speech rhythm. Ambient language influence is evident from the earliest connected speech productions (2 years). Cross-linguistic similarities in child productions are also apparent, when compared with adult speech, reflecting universal challenges in acquiring general articulatory skills. However, though some properties are more characteristic of so-called 'syllable-timing', others are quite the opposite, and this suggests that 'evenness' is not an across-the-board intrinsic property of early productions. The gap between child and adult scores narrows over the age range investigated, and by 6 years the vocalic components of rhythmic profiles are statistically indistinct from those of adult rhythmic profiles for the same language, while the consonantal component is closer, even if still off target.

It should be borne in mind that the differences investigated here are in acoustic indices which, it is claimed, directly reflect a cross-linguistic perceptual distinction in rhythm in adult speech. The perceptual validity of these indexical differences in child speech needs to be tested before we can conclude that what the indices *point to* as rhythmically distinct is, indeed, *perceived* as rhythmically distinct. The same is also true of course for any study of adult speech rhythm and is not a problem unique to child speech. However, even if differences in interval-based indices were not to be perceived as rhythmic in child speech – they could, for example, prove to be weak or incoherent cues perceptually – knowledge of interval variability and vocalic-ness in child speech nevertheless provides insight into when and how the basis of perceived cross-linguistic rhythmic differences in adult speech begins to emerge in early productions.

Other potential dimensions to the perception of rhythm need to be explored. Prieto et al. (submitted) provide evidence that rhythm class distinctions finely correlate with differences in the way languages instantiate the durational marking of prosodic heads and prosodic boundary lengthening. White, Payne, and Mattys (2009) put forward the claim that rhythmic differences between northern and southern varieties of Italian stem from differences in the realization of prosodic phenomena. In addition to an amplified *structural* consideration of the sources of rhythm, other possible differences in the way rhythm is *manifested* need to be investigated, such as intensity,[18] vowel quality and F0. These further considerations lie beyond the scope of this paper, and remain to be explored more fully in both adult and child speech.

Finally, the present study has revealed a non-simplistic picture of how language-dependent rhythmic properties develop in child speech. It is not the case that child speech is categorically more 'even-timed' than adult speech, and it is not necessarily the case that all language-specific rhythmic properties increase with age. The complexity may in part result from an interplay between phonetic and phonological factors, and between general and language-specific factors, in developmental speech. These factors may make competing demands on the child's developing abilities. Furthermore, they need not be synchronized, as the different behavior and different developmental time spans regarding consonant and vocalic interval variability have shown: in much simplified terms, it has been suggested that vocalic interval variability is largely a skill to be mastered, while consonant interval variability is partly an artefact of immature articulatory co-ordination skills (interacting with language-specific phonotactics). Behind each metric may lie multiple phonetic and/or phonological 'causes', and this is applicable to both adult and child speech. As with rhythmic studies of adult speech, a full 'explanation' of the metrics – for example to know whether %V is higher in child speech because of extra phonetic lengthening or because of a higher incidence of open syllables – will also require analysis of the segmental structure and its phonetic implementation. In identifying robust patterns in child speech, and establishing interval-based metrics as a valid tool in the investigation of rhythm acquisition, the findings of this study mark a key step in this direction.

## Acknowledgements

The authors would like to thank Greg Kochanski, Anne Christophe, Volker Dellwo and two anonymous reviewers for helpful comments on an earlier draft of this paper, and Bill Wells, Chris Lee and other colleagues who provided helpful discussions of related presentations at BAAP 2008 and EASR 2008. We are also grateful to Núria Argemí, Anna Barberà, Ana Estrella, and Melanie Bell for recording the data in different languages, and especially to Naomi Hilton for performing the segmentation of the data.

## Funding

This work has been supported by the Generalitat de Catalunya (Batista i Roca grant number 2007 PBR 29, with Pilar Prieto as PI).

## Notes

1 Following White and Mattys (2007), other rhythm metrics, based on variability in the duration of phonological constituents (e.g., Deterding, 2001), were not applied.
2 The term 'intermediate' is used to mean displaying characteristics of both stress- and syllable-timing and therefore, supposedly, falling between these two categories (see Nespor, 1990) along a kind of rhythmic continuum (see Dauer, 1987). As Ramus et al. (1999, p. 269) point out, the existence of languages that fail to fall neatly into these two categories may indicate that there are simply more categories.
3 In a later study, it is argued that Polish may simply belong to a distinct rhythmic class, which is not intermediate between the two others (Ramus et al., 2003, p. 341), and as a consequence the ambiguity over $\Delta$V is potentially resolved.
4 VarcoC also varied according to speakers' intended speech rate within a particular language, though not in a consistent way across the languages in that study (French, German and English).
5 Or %C, since variation in %C and %V amount to the same thing.
6 Elsewhere, however, there is evidence that rate-normalized consonant metrics *do* discriminate. For example, Dellwo (2006) reports a higher VarcoC for English and German than for French (though no statistical tests were reported).
7 These children were the same subjects as used in the present study.
8 Where the concept of 'x-word point' is defined as the point in the child's development at which she can produce x different word types. (See for example Vihman, 1996, p. 141.)
9 The authors do not state the age range for this point, but the range for the 25-word point was 16–17 months.
10 Vowel reduction in English, in the form of centralization to schwa, is of course also qualitative. Qualitative differences – not detectable by duration-based rhythm metrics – may also contribute to the rhythm percept, but are not detectable by duration-based metrics, and lie outside the scope of this study.
11 Although other factors may be present. For example, languages with systematically variable consonant duration (i.e., gemination) should also, in theory, show higher variability in consonant intervals (see White, Payne, & Mattys, 2009, for a comparison of rhythm in geminating and non-geminating varieties of Italian).
12 'More vocalic' or 'less vocalic' are to be understood as referring to the relative proportion of vocalic intervals in an utterance, not the number of vocalic segments.
13 Some adults had more than one child participating, hence there were fewer adults than children recorded. In total, there were 9 English-speaking mothers, 6 Spanish-speaking mothers, and 7 Catalan-speaking mothers.
14 Or an approximation.
15 The correlation between number of syllables and %V was just above the level of significance (Pearson's r for %V $r(1148) = 0.057$, $p = 0.054$).

16 Catalan was not investigated in that study.
17 Given the structural and implementational variability in child speech, both among speakers and longitudinally.
18 In a rare study of this kind, Lee and Todd (2004) report evidence that ‘stress-timed’ English and Dutch show greater variability in vocalic intensity than ‘syllable-timed’ Italian and French, for adult speech.

## References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Allen, G. D., & Hawkins, S. (1978). The development of phonological rhythm. In A. Bell & J. Hooper (Eds.), *Syllables and segments* (pp. 172–185). Amsterdam: North-Holland Publishing.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology. Volume 1: Production* (pp. 227–256). New York: Academic Press.

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.

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.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In Proceedings of the 15th International Congress of Phonetics Sciences, Barcelona, 2693–2696.

Bertinetto, P. M. (1981). *Strutture Prosodiche dell'Italiano*. Firenze: Accademia della Crusca.

Blasdell, R., & Jensen, P. (1970). Stress and word position as determinants of imitation in first-language learners. *Journal of Speech and Hearing Research*, *13*, 193–202.

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/

Boysson-Bardies, B., de, Sagart, L., Halle, P., & Durand, C. (1986). Acoustic investigation of crosslinguistic variability in babbling. In B. Lindblom, & R. Zetterström (Eds.), *Precursors of early speech* (pp. 113–126). New York: Stockton Press.

Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*, 297–319.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research*, *50/4*, 999–1014.

Dasher, R., & Bolinger, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, *12*, 58–71.

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. Proceedings of 11th ICPhS, Tallinn, 447–450.

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, *59*, 75–107.

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

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, Piliscsaba 2003* (pp. 231–241). Frankfurt: Peter Lang.

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. Proceedings of the 15th International Congress of Phonetics Sciences, Barcelona, 471–474.

DePaolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, *36*, 406–422.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*, 217–230.

Ferragne, E., & Pellegrino, F. (2004). A comparative account of the suprasegmental and rhythmic features of British English dialects. Proceedings of "Modelisations pour l'Identification des Langues", Paris, 121–126.

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*, 1208–1211.

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.

Grabe, E., Watson, E., & Post, B. (1999). The acquisition of rhythmic patterns in English and French. *Proceedings ICPhS 1999, San Francisco*, 1201–1204.

Hepper, P. G., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*, 147–153.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*, 262–274.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge: Cambridge University Press.

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *364*, 2325–2335.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630–645.

Kehoe, M., & Lleó, C. (2002). *The emergence of language-specific rhythm in German-Spanish bilingual children*. Paper presented at the Joint Conference of the IX International Congress for the Study of Child Language and the Symposium on Research in Child Language Disorders, Madison, WI.

Kehoe, M., Stoel-Gammon, C. K., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, *38*, 338–350.

Klein, H. (1984). Learning to stress: A case study. *Journal of Child Language*, *11*, 375–390.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*, 2425–2438.

Lea, W. A. (1974). Univac Report PX10791. *Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding*. St. Paul, MN: Sperry Univac.

Lee, C. S., & Todd, N. P. M. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition*, *93*, 225–254.

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

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*, 1994–1997.

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.

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.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*, 1–19.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*, 233–243.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition* (pp. 157–175). Dordrecht: Foris.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 155–176). Amsterdam/Philadelphia: John Benjamins.

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, 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 sulle tipologie ritmiche* (pp. 147–183). Roma: Aracne Biblioteca di Linguistica.

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

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

Pointon, G. E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics*, *8*, 293–304.

Prieto, P., Vanrell, M., Astruc, L., Payne, E., & Post, B. (2008). *Is there any evidence for top-down planning of rhythm? Evidence from Catalan, English and Spanish*. Poster presented at UCL Workshop on Rhythm, March 2008.

Prieto, P., Vanrell, M., Astruc, L., Payne, E., & Post, B. (*submitted*). Phonotactic and phrasal properties of speech rhythm: Evidence from Catalan, English, and Spanish. *Speech Communication*.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crepin, G. (1988). Fetal hearing. *European Journal of Obstetrics and Reproductive Biology*, *28*, 191–212.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. Proceedings of the 15th International Congress of Phonetics Sciences, Barcelona, 337–342.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Risley, T. R., & Reynolds, N. J. (1970). Emphasis as a prompt for verbal imitation. *Journal of Applied Behavior Analysis*, *3*, 185–190.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Edward Arnold.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Galles, N., Alves Limissuri, R., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*, 914–919.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research*, *37*, 831–840.

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development*, *77*, 281–296.

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

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford: Blackwell.

Vihman, M. M., DePaolis, R. A., & Davis, B. L. (1998). Is there a 'Trochaic Bias' in early word learning? Evidence from infant production in English and French. *Child Development*, *69*, 933–947.

Vihman, M. M., Kay, E., Boysson-Bardies, B., Durand, C., & Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology*, *30*, 651–662.

Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory Phonology 8: Phonology and phonetics* (pp. 341–366). New York: Mouton de Gruyter.

Vihman, M. M., & Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*, *57*, 255–266.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth. *Developmental Science*, *10*, 159–164.

White, L., & Mattys, S. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*, 501–522.

White, L., Payne, E., & Mattys, S. (2009). Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigario, S. Frota, & M. Freitas (Eds.), *Interactions in phonetics and phonology* (pp. 137–158). Amsterdam/Philadelphia: Benjamins.

Whitworth, N. (2002). Speech rhythm production in three German–English bilingual families. *Leeds Working Papers in Linguistics and Phonetics*, *9*, 175–205.

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.

## Appendix

**Table 4.** Means (standard errors) of metrics for Spanish, Catalan and English adult speech.

| | Spanish | Catalan | English |
|---|---|---|---|
| %V | 48 (0.7) | 44 (0.6) | 41 (0.6) |
| VarcoV | 50 (2.1) | 42 (1.2) | 61 (1.6) |
| nPVI-V | 44 (1.4) | 40 (1.2) | 62 (1.6) |
| $\Delta$C | 54 (1.6) | 50 (1.7) | 73 (2.9) |
| r*P*VI-C | 57 (2.7) | 53 (1.7) | 72 (2.7) |

**Table 5.** Means (standard errors) of metrics for Spanish, Catalan and English child speech.

| | Spanish | Catalan | English |
|---|---|---|---|
| %V | 51 (0.7) | 51 (0.7) | 47 (0.8) |
| VarcoV | 41 (1.2) | 40 (1.1) | 55 (1.7) |
| nPVI-V | 43 (1.4) | 43 (1.3) | 60 (2.0) |
| $\Delta$C | 83 (3.5) | 77 (2.8) | 112 (4.7) |
| rPVI-C | 88 (3.3) | 88 (3.1) | 126 (6.1) |

**Table 6.** Mean metric scores for child speech by age.

| Metric | Eng child age 2 | Eng child age 4 | Eng child age 6 | Cat child age 2 | Cat child age 4 | Cat child age 6 | Sp child age 2 | Sp child age 4 | Sp child age 6 |
|---|---|---|---|---|---|---|---|---|---|
| $\Delta$C | 145.49 | 102.21 | 91.46 | 75.56 | 89.91 | 66.09 | 80.83 | 90.34 | 75.13 |
| rPVI-C | 178.55 | 101.64 | 93.45 | 92.73 | 89.37 | 75.6 | 92.52 | 89.14 | 81.16 |
| %V | 44.38 | 50.18 | 43.77 | 56.34 | 48.1 | 47.52 | 56.6 | 50.05 | 46.21 |
| VarcoV | 50.57 | 55.84 | 56.62 | 36.82 | 40.59 | 41.32 | 37.06 | 42.07 | 43.13 |
| nPVI-V | 61.36 | 60.93 | 56.47 | 41.23 | 42.53 | 42.6 | 40.45 | 43.19 | 42.84 |