Rhythmic modification in Child Directed Speech[*]

Elinor Payne, Brechtje Post, Lluïsa Astruc, Pilar Prieto, Maria del Mar Vanrell

Interval-based rhythm metrics were applied to the speech of English, Catalan and Spanish female adults addressing their children (aged 2, 4 and 6 years) and compared with their speech when addressing adults. Results reveal that when mothers speak to their children, their speech is more vocalic and more even-timed than when they address other adults. However, cross-linguistic differences are still maintained, suggesting that indices specific to both speech style and language co-exist.

1. Introduction

This paper[1] examines the extent to which the rhythmic properties of adult speech vary according to situational context, and specifically whether the speaker is addressing a child or another adult. To measure rhythmic properties we apply a set of well known 'rhythm metrics' which extract and quantify certain (predominantly durational) properties[2] from the acoustic signal (cf Ramus et al, 1999; Dellwo, 2004; Low et al, 2000; Grabe & Low, 2002; White & Mattys, 2007). While these metrics have been widely applied to capturing perceived cross-linguistic differences in rhythm in (largely non-spontaneous) reading styles of adult speech, little attention has been paid to determining whether such differences are evident in different speech contexts, known to trigger different speaking styles, e.g. child-directed-speech (CDS). We examine the rhythmic properties of CDS for English, Spanish and Catalan, as a function of child age, and compare with adult-directed speech (ADS) in the same languages. We consider the possible interplay of structural and performance factors in the generation of rhythmic percept, and also the degree to which temporal modifications in CDS may be interpreted as accommodations towards the speech of the child in question.

1.1 Background on speech rhythm measurement

Though the acoustic basis of perceived cross-linguistic differences is elusive, the percept of a distinction between two broad language categories (i.e. between the 'Morse Code' rhythm of e.g. Germanic languages and the 'machine gun' rhythm of e.g. Romance languages) is strong, and empirically supported. For example, Ramus, Dupoux and Mehler, 2003, show adult ability to distinguish *between* rhythmic categories, but not *within* them. The validity and efficacy of rhythm metrics in capturing perceived differences have been widely discussed and evaluated (see in particular an overview and comparison of all metrics by White and Mattys, 2007), and only a brief outline of their development will be given here.

Empirical studies (cf. Pointon, 1980; Lea, 1974; Dauer, 1983; Roach, 1982 found no evidence to support the early thesis (Pike, 1945; Abercrombie, 1967) that cross-

[*] This research has been funded by a Batista i Roca research project entitled "The acquisition of rhythm in Catalan, Spanish and English (Ref.: 2007 PBR 29), awarded by the Generalitat de Catalunya, with Pilar Prieto as PI. We are 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.
[1] Many thanks to Greg Kochanski for helpful comments on an earlier draft of this paper, and to Bill Wells, Chris Lee and other colleagues who provided helpful discussion of related presentations at BAAP 2008 and EASR 2008.
[2] Following White and Mattys (2007), other rhythm metrics, based on variability in the duration of phonological constituents (e.g. Deterding, 2001), were not applied.

linguistic differences result from top-down timing strategies selecting different units (foot versus syllable[3]) for isochrony. Following the empirical discreditation of the isochrony account, an alternative explanation (see Bertinetto, 1981; Dauer, 1983; 1987; Roach, 1982; Dasher and Bollinger, 1982) proposed that rhythmic distinction 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, it was observed that in so-called 'stress-timed' languages, like English, there is a greater range of syllable structures (permitting complex codas and onsets), heavier syllables are more likely to attract stress, and unstressed vowels tend to be reduced (both durationally and quantitatively, see Delattre, 1966). In contrast, in so-called 'syllable-timed' languages like Spanish, open syllables are far more common, and vowel reduction is much less evident. This 'phonologically-derived' hypothesis of rhythmic difference in effect captures differences in the way languages execute prosodic prominence. On this interpretation, the execution of prosodic prominence is not dependent (or 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', it is because they have access to the acoustic effects of this interaction, without having to 'know' it.

These claims find strong support in psycholinguistic studies (Nazzi, Bertoncini and Mehler, 1998; see also Ramus et al, 2003; Nazzi & Ramus, 2003; Nazzi et al, 2000) showing that infants attend to rhythmic differences from birth, indicating that they reflect something that can be perceived before linguistic analysis is available to the infant, and therefore can be captured objectively from the acoustic stream. From this, Ramus, Nespor and Mehler (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)." 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 (ΔV); ii) the standard deviation of intervocalic (i.e. consonantal) intervals (ΔC); and the proportion of utterance duration which comprises vocalic intervals (%V). Application of these metrics (Ramus et al, 1999; Ramus et al, 2003) to languages of perceived different rhythmic categories revealed a combination of ΔV and either ΔC or %V to be the most useful in distinguishing categories.

A different approach to measuring the basis of rhythmic difference, but one nevertheless derived from acoustic intervals, uses the pairwise variability index (PVI), proposed by Low et al (2000). 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 the differences between successive intervals, and is normalised (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 normalising 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, normalising eliminates rhythmic difference. Thus, normalising speech variability over an utterance is deemed to be helpful for capturing vocalic interval variability, but unhelpful for capturing consonantal interval variability.

[3] Whence the terms 'stress-timed' and 'syllable-timed'.

The decision as to whether to normalise for speech rate is non-trivial. Evidence that ΔC and ΔV are inversely proportional to speech rate (e.g. Barry, Andreeva, Russo, Dimitrova and Kostadinova, 2003; Dellwo and Wagner, 2003; Lee and Todd, 2004) led Dellwo (2004) to propose the normalised VarcoC and Ferragne and Pellegrino (2004) and White and Mattys (2007) to propose the normalised VarcoV. Indeed, these normalised metrics performed better at distinguishing rhythmic categories than their non-normalised counterparts. Conversely, Dellwo and Wagner (2003) found %V not to vary consistently with speech rate. Thus, there is general consensus that metrics for variability in vocalic intervals should be normalised, but greater uncertainty about the equivalent consonant metrics. However, a potential confound, as noted by White and Mattys (2007), is that variation in speech rate may itself contribute to the rhythmic percept. This may be particularly important when comparing different speech styles within a given language, where the underlying structural properties (presumably) remain constant[4], but parameters of performance, such as speech rate, may be systematically different. In this case, one could argue that the difference is one of rate rather than rhythm. However, if a different rate systematically alters the rhythmic percept, and is systematically used for a given speech style, the boundary is a fuzzy one. Ideally, language-dependent rate differences would be distinguished from individual, style-dependent rate differences. At present, we know of no satisfactory method of distinguishing the two and so are obliged to conflate them in the present study.

With the proliferation of different metrics, the question arises as to how well they perform, comparatively. 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 that %V[5] and the rate-normalised vowel metrics (VarcoV and nPVI-V) to be the most effective (with a slight favouring of VarcoV over nPVI-V, also for within-category discrimination).

Measures of consonantal interval variation proved to be far less effective or consistent in discriminating between rhythmic types. However, as White and Mattys (2007: 18) point out, this may be due to the idiosyncrasies of particular materials, and they cite previous studies (e.g. Ramus et al, 1999) that did achieve greater discriminatory success with consonant interval metrics. Furthermore, as has been noted, the appropriateness of rate normalisation for consonant interval metrics is also questionable. The role of consonant variability in the creation of the percept of rhythmic difference remains, therefore, rather elusive. Further investigation will require careful consideration of the role of speech rate and of the segmental composition of the materials used.

1.2 Characteristics of Child Directed Speech

From a purely descriptive point of view, it is well documented that certain properties of CDS may differ markedly from those observed in Adult Directed Speech (ADS), notably in prosodic patterns and speaking rate (see e.g. Albin and Echols, 1996; Fernald; Fernald & Mazzie 1991; Fernald & Simon 1984; Fernald et al, 1989; Garnica, 1977; Grieser & Kuhl 1988), but also in segmental properties (see e.g. Watt, Docherty, Foulkes, 2003). Common features of CDS that have been identified by these and other studies include higher and greater range of pitch, especially in stressed syllables; longer duration of individual words; prominent final lengthening; slower speech rate; higher amplitude; longer pauses, and more reliable positioning of pauses at phrase boundaries.

[4] There is no reason to predict that syllables or segments are deleted in CDS.
[5] Or %C, since these amount to the same thing.

Some of these modifications, in particular final lengthening, could reasonably be expected to influence rhythm, and therefore to impact on the rhythm metrics. CDS may, however, present other, more generalised, modifications to vocalic and consonantal interval timing, which, one would expect, would also be reflected in the metrics. Any modifications may (and equally may not) be of a systematic nature. If they turned out to be systematic, it could reflect, for example, an accommodation towards temporal properties exhibited in the speech of the child interlocutor, or a particular strategy aimed at bonding or some kind of perceptual facilitation or instruction. One possibility is that adults exaggerate temporal properties characteristic of ADS. Indeed, Dominey and Dodane (2004: 128) claim that 'the essential acoustic property of CDS is the exaggeration or modulation of characteristics that are *already present* in ADS'[6], though they refer to general prosodic characteristics and not to rhythm specifically.

If this exaggeration were true for rhythm, we would predict Spanish CDS to be more even-timed and English CDS to be less even-timed than their ADS equivalents, particularly for speech towards younger infants, and, as a result, for the cross-linguistic differences to be greater for this style of speech. However, cross-linguistic differences in ADS are claimed to emerge from structural differences (at least in part), and such differences are discrete and categorical (syllables are either predominantly open or they are not). It would be difficult to conceive of a way in which such structural properties could be exaggerated. More problematically still, there is simply no reason to expect any structural differences (a difference in syllable structure, for example) between CDS and ADS: unless an adult were mimicking child phonology, why would she delete or simplify consonant clusters? Any observed rhythmic differences, therefore, can be presumed to stem from variation in performance.

From a functional point of view, it has been argued that the richness of prosodic cues in CDS may actively function to attract attention, highlight linguistic structure and aid word identification. In support of this thesis, Kemler Nelson et al (1989) showed that 7- and 10-month old infants were sensitive to prosodic cues to clause boundaries in CDS but not Adult-Directed Speech (ADS). Evidence also suggests (Fernald, 2000; Shockey and Bond, 1980) that properties of CDS evolve as a function of the child's development. Determining whether any CDS modifications are used or indeed are even of any potential linguistic use to children is no trivial task, and it is highly debatable whether this factor could be successfully isolated and therefore whether a truly useful test be conducted. Nevertheless, the application of rhythm metrics to CDS allows us to test for variation between adult speaking styles/contexts.

2 Objectives of present study

The present study aims to determine the extent to which cross-linguistic differences in certain temporal properties (claimed to give rise to specific rhythmic percepts) are detectable also in CDS. Our prediction was that such differences would be present to some degree, given the unlikelihood of structural divergence. It also aims to determine whether any cross-linguistic *similarities* are detectable in CDS, when compared with ADS. Here, our prediction was that differences between CDS and ADS would be observable, but that these would not necessarily be systematic, or similar across languages, given the performative nature of their source.

The languages chosen for this study differ in certain phonological and prosodic properties, and are said to differ according to rhythmic class (for ADS).

[6] Echoed elsewhere in the CDS literature, e.g. Albin and Echols (1996).

i) **English** displays a wide variety of syllable structure types, quantitative vowel reduction (Delattre, 1966), stress attraction to heavier syllables, and substantial final lengthening (Wightman et al, 1992). Rhythmically, it is classically defined as 'stress-timed'.

ii) **Spanish** is dominated by CV syllable structure, and displays weak vowel reduction (Delattre, 1966), a weak correlation between stress and syllable weight, and weaker final lengthening. Rhythmically, it is classically defined as 'syllable-timed.'

iii) **Catalan** displays mostly CV structure (though has a greater incidence of more complex syllables than Spanish), weak vowel reduction, a moderate correlation between stress and syllable weight, and more final lengthening than in Spanish (Ortega-Llebaria & Prieto, 2007). The fact that some of these properties are more typical of 'syllable-timed' languages and some more typical of 'stress-timed' languages has led some researchers to conclude that Catalan is rhythmically 'intermediate' (Nespor, 1990; Ramus et al, 1999). However, there is no firm consensus on its rhythmic status. Prieto, Vanrell, Astruc, Payne and Post (2008) show that when syllable structure properties are controlled for, no durational planning differences arise between Catalan and Spanish. Further evidence for the 'syllable-timed' status of Catalan is given by Gavaldà-Ferré (2007), who shows that different degrees of vowel reduction found in different dialects of Catalan make no difference to rhythm.

2.1 Summary of hypotheses

We hypothesised that:

H1 CDS scores would show some degree of cross-linguistic divergence, as observed in ADS;

H2 CDS may still show rhythmic divergence from ADS, but in an unsystematic way cross-linguistically.

3 Method

3.1 Participants

For each language, we recorded twelve mother-child dialogues in each language (to elicit CDS), and made separate recordings of the mothers interacting with an adult interviewer (to elicit ADS). The children fell into three age groups (i.e. there were four children in each age group for each language): 2-, 4- and 6-year-olds. Some mothers were recorded in two different dialogues, with children of different ages. In total, 36 children and 26 adults were recorded[7].

3.2 Materials and elicitation

The data consisted of short question-and-answer dialogues, elicited through the medium of a structured game, based on short, animated clips, shown on Powerpoint slides

[7] This study was part of a wider study which also investigated rhythm in child speech.

on a laptop screen. The animations showed simple, everyday scenes, which could easily be described in words that were highly familiar to the children. For example, one scene showed a little girl blowing soap bubbles, another showed a little boy playing with building blocks.

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!"*[8]
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).

Table 1 shows the number of utterances produced by mothers in dialogue with each age group, and in dialogue with other adults, for each language. Spanish CDS utterances directed at 6-year olds are missing: this is because, in this particular recording session, the mothers were mistakenly not instructed to repeat what their child had said.

Table 1

	CDS 2-year olds	CDS 4-year olds	CDS 6-year olds	CDS total	ADS
English	69	77	68	214	226
Catalan	69	64	68	201	154
Spanish	65	33	-	98	138

Recordings were made respectively in Cambridge, Madrid and Barcelona, using a Marantz PMD660 recorder and Shure PG81 microphones, in the participants' homes.

3.3 Analysis

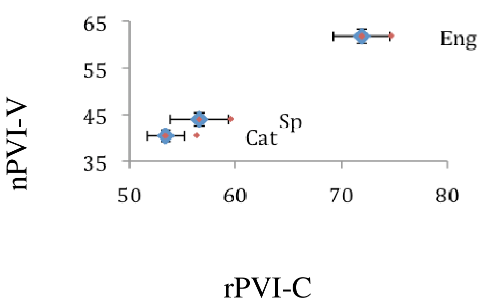
Vocalic intervals were segmented and labelled (start-points and end-points) from the waveform and spectrogram in Praat using standard criteria (e.g. Peterson and Lehiste, 1960). Vocalic and intervocalic (consonantal) intervals were then extracted using a Praat script, and the following metrics were calculated (after Grabe & Low, 2002 and White & Mattys, 2007): %V; ΔV; VarcoV; ΔC; VarcoC; rPVI-V; nPVI-V; nPVI-C and nPVI-C. 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.

Using the SPSS software package, a series of univariate ANOVAS were run separately on the CDS and ADS scores, and a repeated measures analysis was run to compare CDS and ADS scores, with speaking style as the repeated measure. Bonferroni post hoc comparisons were performed on language and age. The p-level is only reported for significant (i.e. $p < .05$) or near-significant differences (i.e. $05 < p \leq .10$).

4 Results and Discussion

[8] Or an approximation.

As a first step, we checked the robustness of the cross-linguistic differences for ADS, to verify whether the elicitation technique used could replicate previous findings for these languages for adult speech. For ADS, 'language' was found to be a main effect for all metrics and for speech rate (see Table 5, Appendix, for the results of statistical tests). However, the post hoc comparisons reveal that the only robust difference is between English and from Spanish and Catalan, while the latter two are more similar. English is distinct from Catalan along all measures, and from Spanish along all measures except nPVI-C. As we would expect for a 'stress-timed' language, English shows greater variability in both consonant and vocalic intervals (see Figures 1a and 1b, and Table 2), and a lower %V (see Figure 1b) and speech rate (Figure 7). 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 reproduce a very similar pattern to the results for Spanish and English in that study[9] (compare with their Figures 1 and 2, p. 511-12).



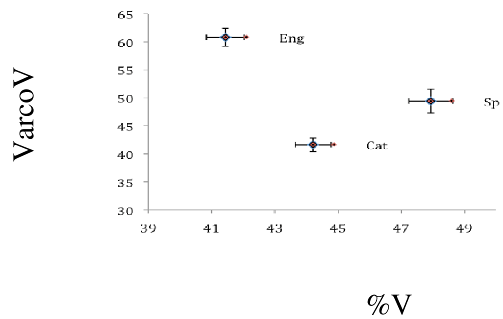


Figure 1a (left): Distribution of Catalan, Spanish and English ADS over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean

Figure 1b (right): Distribution of Catalan, Spanish and English ADS over the %V, VarcoV plane. Bars represent one standard error around the mean

Table 2: Means (standard errors) of rhythm metrics for Spanish, Catalan and English ADS

	Spanish	Catalan	English
Interval measures			
ΔV	48 (3.5)	34 (1.2)	58 (2.1)
ΔC	54 (1.6)	50 (1.7)	73 (2.9)
%V	48 (0.7)	44 (0.6)	41 (0.6)
VarcoV	50 (2.1)	42 (1.2)	61 (1.6)
VarcoC	57 (1.9)	49 (1.2)	63 (1.5)
Pairwise Variability Indices			
nPVI-V	44 (1.4)	40 (1.2)	62 (1.6)
rPVI-V	48 (3.5)	34 (1.3)	60 (2.1)
nPVI-C	61 (1.6)	57 (1.5)	63 (1.5)
rPVI-C	57 (2.7)	53 (1.7)	72 (2.7)
Speech rate			

[9] Catalan was not investigated in that study.

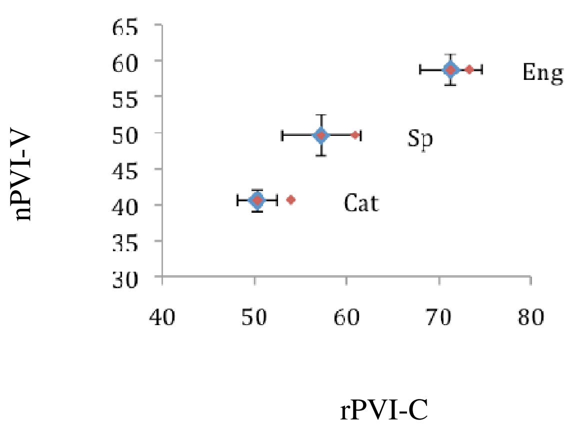
Syllables/second	5.8 (0.1)	5.6 (0.1)	4.6 (0.1)

Some distinction between Catalan and Spanish is detectable from the metrics, but this is significant in the main only for vocalic scores (excepting nPVI-V, see Table 5). For consonant scores, the two languages are only significantly different for VarcoC. %V is significantly lower in Catalan, as would be expected if Catalan does indeed fall in an intermediate rhythmic category. However, for *variability* scores, even where distinctions are detected, Catalan scores are actually *lower* than Spanish, and this is contrary to claims of Catalan being less syllable-timed than Spanish (Nespor, 1990).

4.1 Hypothesis I: Cross-linguistic differences in CDS

Our first hypothesis was that CDS would show the cross-linguistic divergence in rhythmic scores observed in ADS. The results support this hypothesis: as with ADS, language was a main effect for all metrics (see Table 9, Appendix). For all metrics, as with ADS, English shows greater variability in both consonant and vocalic intervals (see Figures 3a and 3b), and a lower %V (Figure 3b) and speech rate (Figure 7). Post hoc comparisons showed that, as for ADS, Catalan is distinct from English in all measures. However, the distinction between the other two languages is made along fewer parameters. English is distinct from Spanish only for rPVIc, nPVIv and rate, while Catalan is distinct from Spanish only in the normalised vocalic metrics (VarcoV and nPVIv). In other words, the associated rhythmic categories for ADS in these languages are maintained to some degree in CDS, although more weakly for the distinction between English and Spanish.

Child age is shown to be a main effect for all consonant metrics, but for none of the vocalic metrics, or for %V or speech rate (see Table 9). Consonant interval variability tends to decrease as child age increases (see Table 6). This is interesting because it has been claimed by some (cf Grabe and Low, 2002) that consonant interval variability is an artefact of a language's phonotactic properties. The fact that adults vary this for different speaking styles in the same language would suggest that this parameter of timing *can* be controlled for speech style. Furthermore, if adults are making fine distinctions in their speech directed at different ages of children, it would appear that control of this parameter is actually quite subtle.



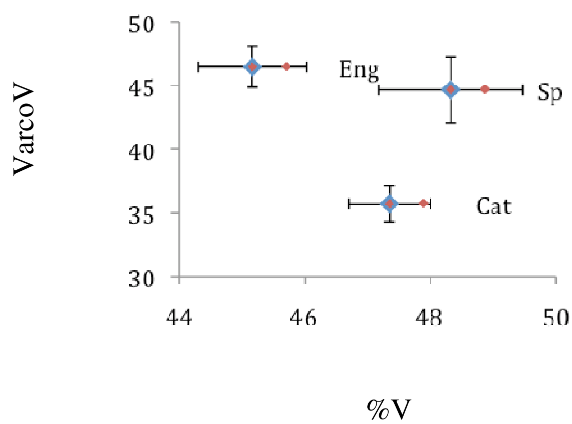


Figure 3a (left): Distribution of Catalan, Spanish and English CDS over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean

Figure 3b (right): Distribution of Catalan, Spanish and English CDS over the %V, VarcoV plane. Bars represent one standard error around the mean

Table 4: Means (standard errors) of rhythm metrics for Spanish, Catalan and English CDS

	Spanish	Catalan	English
Interval measures			
ΔV	44 (3.9)	36 (2.4)	46 (2.1)
ΔC	48 (3.9)	43 (1.5)	55 (2.3)
%V	48 (1.2)	47 (0.7)	45 (0.9)
VarcoV	45 (2.6)	36 (1.4)	47 (1.6)
VarcoC	50 (2.6)	45 (1.4)	54 (1.6)
Pairwise Variability Indices			
nPVI-V	50 (2.9)	41 (1.5)	59 (2.1)
rPVI-V	51 (4.7)	42 (2.4)	58 (2.1)
nPVI-C	62 (3.2)	53 (2.0)	68 (2.1)
rPVI-C	57 (4.3)	50 (2.2)	71 (3.4)
Speech rate			
Syllables/second	5.8 (0.2)	5.5 (0.1)	5.1 (0.1)

4.3 Hypothesis 2: rhythmic properties of CDS

Our second hypothesis, relating to rhythmic type, was that CDS scores would differ, cross-linguistically, from ADS scores, and specifically that they would exaggerate the rhythmic properties of ADS in the language concerned.

4.3.1 %V

%V in CDS proved to be (significantly) higher than in ADS.

4.3.2 Vocalic interval variability

Comparing ADS and CDS, for vocalic interval variability, we find that adult speech style had a significant main effect when variability was measured globally (ΔV and VarcoV), with variability significantly *lower* in CDS than in ADS in all languages (see Figures 4-6b for VarcoV, and Table 10 for statistical results). For sequential measures (PVI), however, there was no significant difference between CDS and ADS (see Figures 4a, 5a and 6a for nPVI-V).

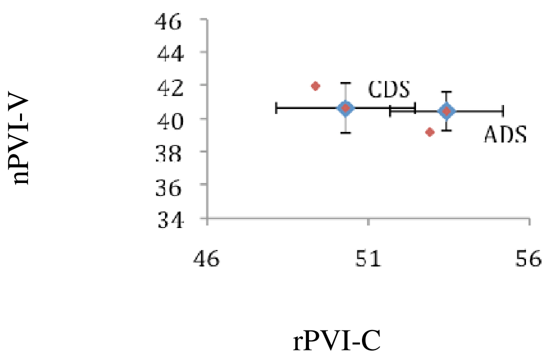
4.3.3 Consonant interval variability

Comparing adult speech styles, consonant interval variability was significantly *lower* in CDS than in ADS (see Table 10, and Figures 4b-6b), when measured globally (i.e. ΔC and

VarcoC). For sequential measures (PVI), however, there was no significant difference between CDS and ADS (see Figures 4-6a, for rPVI-C). These mirror the results for vocalic interval variability.

Figure 4a (left): Distribution of CDS and ADS in Catalan over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean

Figure 4b (right): Distribution of CDS and ADS in Catalan over the %V, VarcoV plane. Bars represent one standard error around the mean



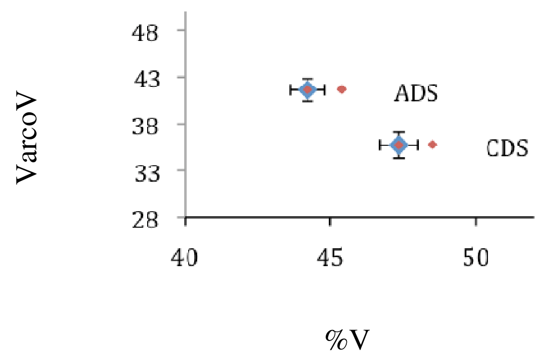
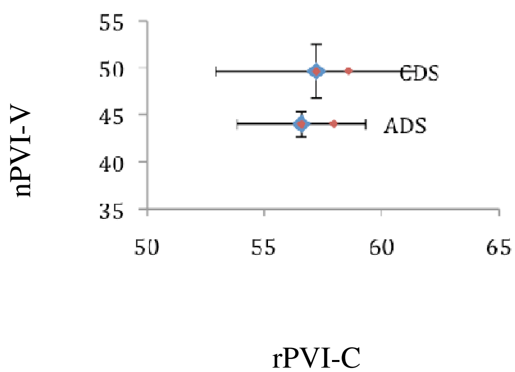


Figure 5a (left): Distribution of CDS and ADS in Spanish over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean

Figure 5b (right): Distribution of CDS and ADS in Spanish over the %V, VarcoV plane. Bars represent one standard error around the mean



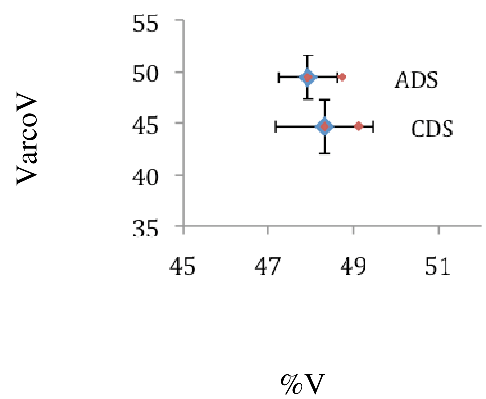
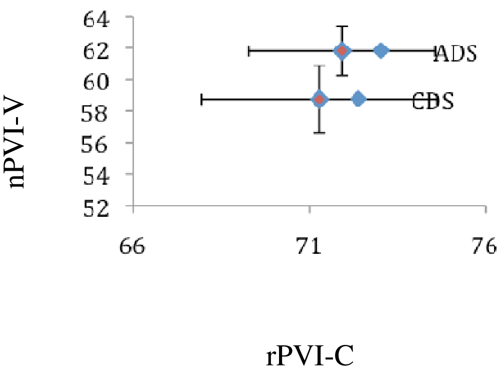
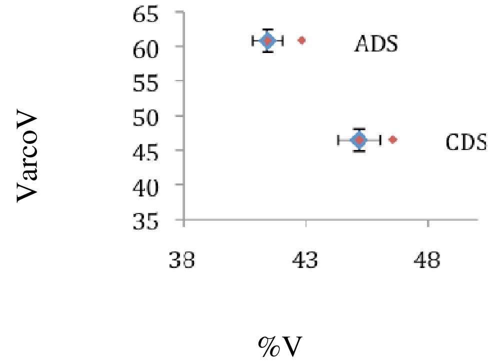


Figure 6a (left): Distribution of CDS and ADS in English over the rPVI-C, nPVI-V plane. Bars represent one standard error around the mean

Figure 6b (right): Distribution of CDS and ADS in English over the %V, VarcoV plane. Bars represent one standard error around the mean





4.3.4 Summary of CDS rhythmic characteristics

For adult speech style, the metrics reveal that, cross-linguistically, mothers showed less *global* variability in both consonant and vowel interval duration, and had more vocalic speech when speaking to their children than when speaking to the adult interviewer. (With the sequential variability (PVI) measures, the pattern is less clear). Since these modifications occurred regardless of the language spoken, they also can be presumed to have occurred regardless of claimed rhythmic class. Higher %V and lower variability are supposedly both characteristic of 'syllable-timing' (more even timing), and therefore it would appear to be the case that generalised rhythmic modifications take place in CDS.

4.4 Overall summary

A clear finding of this study is that rhythmic class distinctions detectable in ADS are also clearly observable in CDS, though in some cases not to the same degree. The differences between English and Catalan are particularly robust: CDS in English has a higher variability and lower %V, (i.e. more 'stress-timed' rhythm) than in other languages. There is a weaker distinction between English and Spanish, and a very weak distinction between Catalan and Spanish. Contrary to some claims (e.g. Nespor, 1990), there is little evidence from these metrics, for either ADS or CDS, to support the view that Catalan should be classed as intermediate between 'stress-timed' and 'syllable-timed': although Catalan has a higher %V than Spanish, it has *lower* variability in both consonant and vowel interval duration, and, if anything, is 'further away' from English.

Cross-linguistically, mothers increase the vocalic proportion and decrease the interval variability of their speech when they address infants (and, in English, somewhat intriguingly, also speak more quickly, especially to older children). This modification results in a slightly weaker distinction between languages for this style of speech (as reported above). For some parameters at least, how mothers modify their speech depends in part on the age of the child they are addressing: this is particularly evident for consonant interval variability, which is already lower than in ADS, but also decreases as the age of the child increases. Higher %V and lower variability are both purportedly characteristics of so-called 'syllable-timing' rhythm, suggesting that adults in all three languages modify their speech towards this end of a hypothesised rhythm continuum (though at different points along this continuum, since the language distinction is maintained). By doing so, the mothers investigated in the study were possibly accommodating towards aspects of the speech of their child (by increasing %V and

lowering vocalic variability, see Payne et al, *submitted*), at the same time as exaggerating the lower consonant variability of normal adult speech.

6 Conclusions and next steps

The interval-based metrics applied in this study yield a clear picture of cross-linguistic rhythmic distinctions (previously established for ADS) co-existing with cross-linguistic similarities in CDS, which is more categorically even-timed than speech directed to other adults. It should be borne in mind that the differences investigated here are in acoustic indices which, it is claimed, *form the basis* of a cross-linguistic perceptual distinction in ADS. The perceptual validity of these differences in CDS needs to be tested before we can conclude that they are, in fact, *rhythmic* in nature also in these kinds of speech. Furthermore, other potential dimensions to the perception of rhythm, such as intensity, vowel quality and F0, lie beyond the scope of this paper, but remain to be explored more fully[10]. Differences in interval-based indices might not to be perceived as rhythmic in CDS – they could, for example, prove perceptually weak or incoherent.

The fact that, from the production perspective, temporal characteristics deemed at least to *contribute* to the percept of rhythm, vary systematically between two styles of adult speech within a given language, shows that the percept of rhythm does not emerge from structural properties alone. It would seem, therefore, that a full model of speech rhythm should be able to account for and incorporate a performance dimension (i.e. phonetic implementation). It is reasonable to conjecture that such a dimension would accommodate also an individual, speaker-specific parameter, in addition to the style-specific (but speaker-shared) parameters as investigated in the present study. Additionally, it would be interesting to examine to degree to which speaker-shared performative characteristics are, at the same time, language-specific; in other words, are there linguistic-phonetic factors which contribute to the percept of speech rhythm? That the speech context investigated here is shown to exhibit a degree of similarity cross-linguistically is of particular interest, since it implies that cross-linguistic performance patterns can and do exist, and would suggest the possibility that manipulation of rhythm has a more general functional purpose (or at least aim).

References

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

ALBIN, D. D., & ECHOLS, C. H. (1996). Characteristics of stressed and word-final syllables in infant-directed speech: Implications for word-level segmentation. Infant Behavior and Development, 19, 401-418.

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 (pp. 2693–2696). Barcelona.

BERNHARDT, B. H. & STEMBERGER, J. P. (1998). Handbook of phonological development from the perspective of constraint–based nonlinear phonology. San Diego: Academic Press.

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

[10] 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 ADS.

BOYSSON-BARDIES, B. de, SAGART, L. HALLE, P. & DURAND, C. (1986). "Acoustic investigation of crosslinguistic variability in babbling." In Lindblom, B. & Zetterström, R. (Eds.), Precursors of early speech. New York: Stockton Press.

BOYSSON-BARDIES, B. and 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, 58–69.

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

DAUER, R. M. (1987). Phonetic and phonological components of language rhythm. In Proceedings of 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. (2004). Rhythm and speech rate: A variation coefficient for deltaC. Proceedings of the 38th Linguistic Colloquium, Budapest.

DELLWO, V., & WAGNER, P. (2003). Relations between language rhythm and speech rate. In Proceedings of the 15th international congress of phonetics sciences (pp. 471–474). Barcelona.

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

DOMINEY, P. F. and DODANE, C. (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. Journal of Neurolinguistics, 17, 121-145.

FERNALD, A. (2000). Speech to Infants as Hyperspeech: Knowledge-Driven Processes in Early Word Recognition. Phonetica 57: 242-254.

FERNALD, A. & MAZZIE, C. (1991). Prosody and focus in speech to infants and adults. Developmental Psychology, 27(2), 209-221.

FERNLAD, A. & SIMON, T. (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20(1), 104-113.

FERRAGNE, E. & PELLEGRINO, F. (2004). A comparative account of the suprasegmental and rhythmic features of British English dialects. In Proceedings of ''Modelisations pour l'Identification des Langues'' Paris.

GARNICA, O., (1977). Some prosodic and paralinguistic features of speech to young children. In Snow, C.E. and Ferguson, C.A., (Eds.), Talking to children: Language input and acquisition, (pp. 63-88), Cambridge, MA: Cambridge University Press.

GAVALDÀ-FERRÈ, N. (2007). Vowel reduction and Catalan speech rhythm. Unpublished MA Thesis, University College London.

GRABE, E. and 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, I. and Post, B. (1999). The acquisition of rhythmic patterns in English and French. In Proceedings ICPhS 1999, San Francisco, (pp. 1201-1204).

GRIESER, D. L., & KUHL, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental Psychology, 24(1), 14-20.

HEPPER, P. G., SCOTT, D. and SHAHIDULLAH, S. (1993). Newborn and fetal response to maternal voice. Journal of Reproductive and Infant Psychology, 11, 147-153.

JUSCZYK, P. W., CUTLER, A. and 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., & CHARLESE-LUCE, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language, 33, 630-645.

KEMLER NELSON, D.G., HIRSH-PASEK, K., JUSCZYK, P.W., and CASSIDY, K.W. (1989). How the prosodic cues in motherese might assist language learning. Journal of Child Language, 16, 55-68.

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

LEE, C. S. and 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.

LLEÓ, C., & KEHOE, M. (2002). On the interaction of phonological systems in child bilingual acquisition. International Journal of Bilingualism, 6, 233–237.

LOW, E.L., GRABE, E. & NOLAN, F. (2000). Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English. Language and Speech, 43, 377-401.

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, M. & PRIETO, P. (2007). Disentangling stress from accent in Spanish: production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró, and M.-J. Solé (ed.) Segmental and Prosodic Issues in Romance Phonology, Amsterdam/Philadelphia, John Benjamins.

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.

QUERLEU, D., RENARD, X., VERSYP, F., PARIS-DELRUE, L. and CREPIN, G. (1988). Fetal hearing. European Journal of Obstetrics and Reproductive Biology 29: 191-212.

RAMUS, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In Proceedings of speech prosody 2002 (pp. 115–120). Aix-en-Provence.

RAMUS, F., DUPOUX, E., & MEHLER, J. (2003). The psychological reality of rhythm classes: Perceptual studies. Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, (pp. 337-342).

RAMUS, F., NESPOR, M., & MEHLER, J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition, 73, 265-292.

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

SAFFRAN, J.R., NWEPORT, E.L., & ASLIN, R.N. (1996). Word segmentation: The role of distributional cues. Journal of Memory and Language, 35, 606-621.

SHOCKEY, L. and BOND, Z. S. (1980). Phonological processes in speech addressed to children. Phonetica, vol. 37, 267-274.

VIHMAN, M. M. (1996). Phonological Development: The Origins of Language in the Child, Oxford: Blackwell.

VIHMAN, M. M., KAY, E., BOYSSON-BARDIES, B., DURAND, C. and 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., 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., 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.

WATT, D., G. DOCHERTY and P. FOULKES (2003). First accent acquisition: a study of phonetic variation in child-directed speech. In Proceedings of the 15th ICPhS, Barcelona, (pp. 1959-1962).

WHITE L. and MATTYS, S. (2007) Calibrating rhythm: First language and second language studies. Journal of Phonetics, 35, 501-522.

WHITE, L., PAYNE, E. and MATTYS, S. (*forthcoming*) Rhythmic and prosodic contrast in Venetan and Sicilian Italian. In M. Vigario, S. Frota & M. Freitas (Eds), Interactions in Phonetics and Phonology, Benjamins.

WHITWORTH, N. (2002). Speech rhythm production in three German–English bilingual families. In 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 1: Inter-language differences in scores for ADS

metric	Language as main effect	En v Ca *p*=	En v Sp *p*=	Ca v Sp *p*=
%V	F(2) = 26.694 p = 0.000	0.004	0.000	0.000
ΔC -	F(2) = 24.850 p = 0.000	0.000	0.000	*1.000*
VarcoC	F(2) = 20.318 p = 0.000	0.000	0.037	0.002
rPVIc	F(2) = 16.913 p = 0.000	0.000	0.000	*1.000*
nPVIc	F(2)= 3.812 p = 0.023	0.018	1.000	*0.308*
ΔV	F(2) = 29.240 p = 0.000	0.000	0.008	0.000
VarcoV	F(2) = 36.079 p = 0.000	0.000	0.000	0.008
rPVIv	F(2) = 32.168 p = 0.000	0.000	0.002	0.000
nPVIv	F(2) = 68.308 p = 0.000	0.000	0.000	*0.336*
Rate (syll/sec)	F(2) = 69.892 p = 0.000	0.000	0.000	0.376

Table 2: Results of ANOVA on CHILD scores

Metric	Language	Age	Age*lang	Post hoc: language	Post hoc: age

%V	F(2)14.355; p=0.000	F(2)17.410; p=0.000	F(4)=11.093; p=0.000	C v E: 0.000 S v E: 0.000	2 v 4: 0.005 2 v 6: 0.000 4 v 6: 0.004
ΔC	F(2)=27.716; p=0.000	F(2)=9.868; p=0.000	F(4)=7.147; p=0.000	C v E: 0.000 S v E: 0.000	2 v 6: 0.000 4 v 6: 0.003
VarcoC	F(2)=20.3; p=0.000	F(2)=3.5; p=0.031	F(4)=2.435; p=0.046	C v E: 0.000 S v E: 0.000 S v C: 0.029	2 v 6: 0.017
rPVIc	F(2)=28.523; p=0.000	F(2)=20.708; p=0.000	F(4)=11.472; p=0.000	C v E: 0.000 S v E: 0.000	2 v 4: 0.000 2 v 6: 0.000 4 v 6: 0.059
nPVIc	F(2)=10.686; p=0.000	F(2)13.546; p=0.000	F(4)=4.684; p=0.001	C v E: 0.000 S v E: 0.012	2 v 4: 0.000 2 v 6: 0.000
ΔV	F(2)=24.887; p=0.000	F(2)=6.410; p=0.002	F(4)=1.617; p=0.168	C v E: 0.000 S v E: 0.000	2 v 6: 0.005 4 v 6: 0.004
VarcoV	F(2)=24.879; p-0.000	F(2)=2.633; p=0.073	F(4)=0.436; p=0.782	C v E: 0.000 S v E: 0.000	2 v 4: 0.049
rPVIv	F(2)=26.222; p=0.000	F(2)10.297; p=0.000	F(4)=2.309; p=0.057	C v E: 0.000 S v E: 0.000	2 v 6: 0.000 4 v 6: 0.003
nPVIv	F(2)=38.299; p=0.000	F(2)=0.297; p=0.743	F(4)=0.622; p=0.647	C v E: 0.000 S v E: 0.000	
Rate	F(2)=33.065; p=0.000	F(2)=43.718; p=0.000	F(4)=1.161; p=0.327	C v E: 0.000 S v E: 0.000	2 v 4: 0.006 2 v 6: 0.000 4 v 6: 0.000

Table 3: Mean consonant interval metric scores for ADS, CDS and CHILD scores (ADS scores cannot be categorised by age).

metric	Age	En ADS	En CDS	En Child	Cat ADS	Cat CDS	Cat Child	Sp ADS	Sp CDS	Sp Child
ΔC	2		62.25	145.49		45.07	75.56		46.32	80.83
	4	73.42	57.17	102.21	49.59	44.26	89.91	53.6	49.87	90.34
	6		44.37	91.46		40.14	66.09			75.13
VarcoC	2		54.5	72.93		50.19	54.64		51.16	54.85
	4	62.5	56.86	63.32	49.11	46.33	52.82	57.04	47.71	59.3
	6		50.85	60.31		38.67	47.6			56.87
rPVI-C	2		81.86	178.55		56.43	89.71		56.93	92.52
	4	71.93	74.43	101.64	53.41	48.03	93.86	56.59	57.78	89.14
	6		57.0	93.45		46.16	69.6			81.16
nPVI-C	2		69.0	89.22		62.48	65.63		64.06	65.21
	4	62.88	70.83	65.25	57.1	51.45	59.66	60.95	58.27	62.66
	6		64.28	61.39		46.06	53.19			63.78

Table 4: Mean vocalic interval metric scores for ADS, CDS and CHILD speech (ADS scores are not categorised by age).

metric	Age	En ADS	En CDS	En Child	Cat ADS	Cat CDS	Cat Child	Sp ADS	Sp CDS	Sp Child
%V	2		42.95	44.38		48.39	56.34		48.7	56.6
	4	41.44	46.53	50.18	44.21	46.72	48.1	47.93	47.56	50.05
	6		45.61	44.66		47.02	48.52			46.21

ΔV	2		44.5	99.76		44.91	67.96		43.59	67.18
	4	58.03	48.84	103.85	33.84	31.32	61.39	48.18	46.11	63.18
	6		43.61	72.85		32.19	55.4			57.73
VarcoV	2		43.59	50.57		39.25	37.82		43.5	36.06
	4	60.82	47.86	53.84	41.64	36.82	40.59	49.45	47.00	42.07
	6		47.42	51.62		31.58	48.52			43.13
rPVI-V	2		54.59	120.99		50.92	79.77		49.11	74.7
	4	59.47	62.35	114.86	33.49	40.03	64.55	47.81	55.85	68.08
	6		55.35	75.84		35.99	60.31			60.69
nPVI-V	2		54.09	61.36		42.61	44.68		47.25	40.45
	4	61.81	61.35	60.93	40.43	44.32	41.63	44.02	54.36	43.19
	6		59.91	56.47		35.31	42.3			42.84

Table 5: Results of ANOVA for CDS scores

Metric	language	Post hoc lang	age	Post hoc age
%V	F(2)=3.413; p=0.034	-	F(2)=0.032; p=0.968	-
ΔC	F(2)=7.867; p=0.000	C v E: 0.000	F(2)=4.816; p=0.008	2 v 6: 0.024 4 v 6: 0.038
VarcoC	F(2)=9.012; p=0.000	C v E: 0.000	F(2)=5.043; 0.007	2 v 6: 0.011 4 v 6: 0.033
rPVIc	F(2)=14.898; p=0.000	C v E: 0.000 S v E: 0.015	F(2)=5.704; 0.004	2 v 6: 0.008
nPVIc	F(2)=12.487; p=0.000	C v E: 0.000	F(2)=4.597; 0.011	2 v 6: 0.009
ΔV	F(2)=4.171; p=0.016	C v E: 0.008	F(2)=0.964; p=0.382	-
VarcoV	F(2)=11.622; p=0.000	C v E: 0.000 C v S: 0.004	F(2)=0.813; p=0.444	-
rPVIv	F(2)=7.382; p=0.001	C v E: 0.000	F(2)=1.206; p=0.300	-
nPVIv	F(2)=22.025; p=0.000	C v E: 0.000; C v S: 0.020; E v S: 0.017	F(2)=2.340; p=0.097	-
Rate	F(2)=9.418; p=0.000	C v E: 0.020 E v S: 0.000	F(2)=2.487; p=0.084	-

Table 6: Results of Anovas for CHILD versus CDS and for CDS versus ADS

metric	CHILD versus CDS	ADS versus CDS
%V	F(1)=20.863; p=0.000	F(1)=6.375; p=0.012
ΔC	F(1)=217.620; p=0.000	F(1)=14.043; p=0.000
VarcoC	F(1)=36.454; p=0.000	F(1)=13.265; p=0.000
rPVIc	F(1)=133.651; p=0.000	F(1)=0.017; p=0.895
nPVIc	F(1)=4.788; p=0.029	F(1)=0.157; p=0.692
ΔV	F(1)=116.061; p=0.000	F(1)=4.626; p=0.032
VarcoV	F(1)=1.721; p=0.190	F(1)=25.670; p=0.000
rPVIv	F(1)=81.357; p=0.000	F(1)=1.251; p=0.264
nPVIv	F(1)=0.934; p=0.334	F(1)=0.436; p=0.509
Rate	F(1)=848.722; p=0.000	F(1)=0.989; p=0.321