Tonal Targets in Early Child English, Spanish, and Catalan

Lluïsa Astruc  
The Open University, UK;  
University of Cambridge, UK

Elinor Payne  
University of Oxford, UK

Brechtje Post  
University of Cambridge, UK

Maria del Mar Vanrell  
Universitat Pompeu Fabra, Spain

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

Abstract  
This study analyses the scaling and alignment of low and high intonational targets in the speech of 27 children – nine English-speaking, nine Catalan-speaking and nine Spanish-speaking – between the ages of two and six years. We compared the intonational patterns of words controlled for number of syllables and stress position in the child speech to the adult target speech provided by their mothers, and to a dataset of adult-directed speech recorded at a later stage for the purpose of measuring pitch height. A corpus of 624 utterances was elicited using a controlled naming task and analysed within the Autosegmental Metrical framework. Measuring the pitch height and pitch timing of nuclear pitch accents, we found that once the effects of syllable duration are accounted for, young children reach tonal targets with remarkable precision. Overall, the results indicate that the phonetic aspects of intonation are acquired from a very early age. Even the youngest children show adult-like alignment of the low target, although mastery of the high target increases with age. Young Spanish-speaking children, however, show a more precise attainment of pitch.
scaling and alignment of their (high) tonal targets than do Catalan and English children; where the ambient language lies within a general prosodic typology appears to influence the acquisition of tonal targets.

**Keywords**
Catalan, child language, English, Spanish, speech development

**Introduction**

While perceptual data on intonational development, especially from the period from birth up to the first year, are relatively abundant (see the review in Vihman, Nakai, & DePaolis, 2006), production data from the early childhood years are scarce. To our knowledge, this is the first large-scale study of tonal alignment and scaling in early child speech. Previous studies typically examine longitudinal data from a reduced sample (often just one child) and also tend to perform contour-based analyses of intonation, and as a result such studies examine the overall shape and direction of the contours rather than the phonetic detail of the different tonal targets (see the review in Snow & Balog, 2002). For instance, these studies may concentrate on one or two parameters, such as the average pitch range over the utterance (Crystal, 1979; Marcos, 1987; Snow, 1995) or the direction of nuclear tones (Galligan, 1986; Snow, 2006; Whalen, Levitt, & Wang, 1991). This approach has yielded useful insights, among them the finding that children have a much more reduced pitch range than adults, and that, at least in English, they acquire falling tones first and final rising tones much later (Snow, 1998; Wells, Peppé, & Goulardris, 2004). Likewise, it has been shown that there are cross-linguistic differences in the type of contours (falling or rising) that are first acquired by the child (e.g. Hallé, de Boysson-Bardies, & Vihman, 1991; Whalen et al., 1991).

Although contour-based analyses have proven useful for describing the intonation of babbling and perhaps also of one-word utterances, they fall short when it comes to describing the intonation of multi-word utterances, as they do not allow a fine-grained study of the phonetic detail of the alignment and scaling of tonal targets with respect to the segmental structure. Thus, it becomes impossible to draw meaningful comparisons to adult data, which are generally analysed following the conventions of the Autosegmental Metrical (henceforth AM) framework.

The AM framework has been applied to many languages and has quickly become the most widely used phonological framework for analysing intonation because it is simple, versatile and allows for cross-linguistic comparison. The AM model (Beckman & Pierrehumbert, 1986; Gussenhoven, 2005; Ladd, 1996; Pierrehumbert, 1984; Silverman et al., 1992) specifies both a phonological and a phonetic component. The inventory of pitch accents and boundary tones constitutes the phonological component (the intonational phonology of the language), and a further phonetic component applies a set of rules, which are language-specific to some extent, that realise the phonological tones as tonal targets. The phonetic component thus specifies the exact location of tones in relation to the segmental string (alignment) and pitch height (scaling). Beyond the application of the phonetic rules, overall pitch height can be further influenced by paralinguistic factors, such as the speaker's psychological condition and communicative purpose (Gussenhoven, 2005, p. 124).

According to the tenets of the AM framework, the intonation of any language can be analysed as a succession of high and low tonal events associated (aligned) either with those syllables that are metrically prominent (e.g. *ba* in *baby*) or with the right-most boundary of phrases. The first set are called pitch accents, or simply accents, and can be high (H*), low (L*) or a combination of high and low. The latter set are called boundary tones and they also can be high (H%), low (L%) or, in
the models that permit it, some combination of high and low (e.g. L-H%). The nuclear accent is the most prominent pitch accent in a phrase, and is also frequently the last one. Prenuclear accents are all those pitch accents preceding the nuclear accent in any given phrase.

Some recent studies that have successfully applied the AM framework to the description of early child intonation are Chen and Fikkert (2007) for Dutch; Frota and Vigário (2008) for Portuguese; Prieto and Vanrell (2007) and Prieto, Estrella, Thorson, and Vanrell (2012) for Catalan. Chen and Fikkert (2007) examined the productions of three children between the ages of 1;4 and 2;1. They found that the children had acquired the basic inventory of nuclear pitch accents and boundary tones by the time they had a vocabulary of about 160 words and the inventory of prenuclear accents by the time they had a vocabulary of 230 words. They argued that there is a correlation between the expansion of the vocabulary and the expansion of the tonal inventory. Frota and Vigário (2008) studied the longitudinal data elicited from a European Portuguese infant recorded at periodic intervals between the ages of 0;9 and 2;6. A preliminary analysis of the data up to the age of 1;9 shows that at this stage the child has acquired the adult inventory of tonal contours and has no major difficulties with their realisation. For instance, while at previous stages in HL* contours the child pushed the end of the fall to the end of the word and realised the L at a higher level than it should be, by 1;9 the alignment and scaling of these tonal targets are essentially correct. Prieto and Vanrell (2007) and Prieto, Estrella et al. (2012) found that the four Catalan children and the two Spanish children in their study displayed a developed intonational grammar by 1;9 and were able to control the phonetic production of a variety of phonologically distinct pitch accents and boundary tones.

The few instrumental studies available so far seem to lend support to the widely accepted assumption that intonation develops more quickly than the timing aspects of prosody. Indeed, it is generally believed that young children show good control of pitch from a very young age, although they still produce excessively long syllables and need time to learn to produce appropriate contrasts in syllable length according to metrical factors such as prominence and position in the word and utterance (Allen & Hawkins, 1978, 1980; Kehoe, Stoel-Gammon, & Buder, 1995; Payne, Post, Astruc, Prieto, & Vanrell, 2012; Snow, 1994). Other studies, however, have found substantial variability in the pitch control of infants and very young children (see DePaolis, Vihman, & Kunnari, 2008).

In line with the tenets of the AM approach, we assume that acquiring the intonational system of a language involves acquiring the phonological inventory plus a set of phonetic rules for mapping the phonological elements onto the segmental string (although acquiring intonation in its totality involves much more than that; for example, it requires knowing the meaning and uses of the tonal patterns, which is to say its semantics and pragmatics). In this study we concern ourselves with the detailed quantitative study of the acquisition of the phonetic implementation rules that govern the alignment and scaling of tonal targets, aspects that have already been extensively studied in adult language (see a summary in Ladd, 2008, Chapter 5). From the studies on adult intonation we know that tones are very precisely aligned with segmental landmarks. Detailed phonetic analyses reveal a consistent tendency for peaks, especially nuclear peaks, to be precisely timed with segmental landmarks such as the syllable onset or offset (Arvaniti, Ladd, & Mennen, 1998, for Greek; Atterer & Ladd, 2004, for German; Ladd, Faulkner, Faulkner, & Schepman, 1999, for English). Such consistent regularities have led researchers to propose the so-called segmental anchoring hypothesis, which says that the beginning and/or ending of tones coincide with specific points in the segmental string. (For a summary of the work on tonal alignment see Ladd, 2008, p. 172; for an approach that emphasises the relevance of the syllabic onset see Xu, 1998, for Mandarin; Xu & Liu, 2006, for Mandarin and English.)

Further research has offered evidence that alignment is sensitive to duration, as determined by speech rate and segmental composition (e.g. syllables differ in the number and type of segments;
segments themselves have different intrinsic durations). This has been proposed for English (Ladd, Schepman, White, Quarmby, & Stackhouse, 2009; van Santen & Hirschberg, 1994); Dutch (Rietveld & Gussenhoven, 1995); Pisa and Bari Italian (Gili-Fivela & Savino, 2003); French (Post, 2011; Welby & Loevenbruck, 2005); and Spanish (Prieto & Torreira, 2007), among others.

Data regarding the alignment of the prenuclear peaks and low targets are far less conclusive. Prenuclear peaks are known to be much more variable than nuclear peaks. As for low tonal targets in English, Pierrehumbert’s original position (e.g. Pierrehumbert, 1984) was to treat valleys (low points between peaks) as *sagging transitions*, that is, mere interpolations between two tonal targets. Ladd (1996, p. 105) reviews the arguments for and against this claim and concludes that this interpretation causes more problems than it resolves, because it complicates the phonetic interpretation of the low tone in other cases, such as the L in L+H* and the optional onglide in H*, as we explain in more detail in Section 3.3. The alignment of L in other languages such as Greek (Arvaniti et al., 1998; Ladd, Mennen, & Schepman, 2000), however, proved very robust. From studies on adult alignment we know that peaks and also perhaps L targets, in the absence of tonal pressure effects (that is, no adjacent pitch accents or boundary tones), tend to be consistently aligned with segmental or syllabic landmarks, although they may be influenced by upcoming prosodic structure (syllable, word, intonational phrase (IP)) and/or speech rate.

2 General goals and research questions

In this study we examine the scaling and tonal alignment of the valley and peak in a contour, (L) H* L% (see the description in Section 3.3), which is comparable across the three languages and is sufficiently abundant in our corpus for a quantitative study.

We ask the following research questions: (1) Are there developmental differences in the phonetic implementation of peak scaling and alignment?; (2) If so, are developmental differences also to be found cross-linguistically?

A positive response to (1) would be in line with the claim that children have a reduced pitch range compared to adults (e.g. Marcos, 1987), and that range increases as the child acquires a complex tonal inventory. However, studies that examine falling and rising tones (e.g. Snow, 1995, 1998) have shown that falling tones behave very differently from rising tones. The four-year-old children in Snow (1998) actually use a pitch range with falling tones that is wider than that of adults. Therefore, we expect the younger children to use a wider pitch range than older children and than adults for the falling tones. We expect, as well, to find developmental effects on alignment, with young children having less precise alignment than older children or adults. That is, both for scaling and alignment, we expect the distance between child and adult to decrease as the child develops linguistically.

With regard to (2), we can expect to find cross-linguistic differences in the age of acquisition of alignment, since English, Catalan and Spanish differ with respect to typological factors such as rhythm, syllable composition and density of pitch accents – factors that are known to interact with the alignment of tonal targets. Previous studies of rhythm in adult speech have examined the possibility that languages might fall into loose clusters, along a scale of more or less even timing and other prominence properties, at least with regard to the percept of rhythm (e.g. Ramus, Nespor, Mehler, 1999). English would appear towards one end of this scale (conventionally described as ‘stress-timing’, and containing complex consonant clusters, vowel reduction, and sparse pitch accents), Spanish towards the other end (conventionally described as ‘syllable-timing’, and containing a predominance of CV syllables and densely occurring pitch accents). Catalan was claimed to fall somewhere between the two, on account of a mixture of properties, but leaning more towards
the ‘syllable-timed’ cluster. More recently, Prieto et al. (2010, 2012), have shown that cross-linguistic differences in the way languages use duration to mark boundaries and prosodic heads contribute rhythmic differences as captured by certain metrics. Using the same adult participants as in the present study and a very controlled methodology, they partially corroborated previous findings, as they found that Spanish and Catalan are more similar to each other than they are to English, including in an analysis of prosodic head and boundary marking.

The few existing cross-linguistic studies of rhythm in children (DePaolis et al., 2008; Grabe, Post, & Watson, 1999; Payne et al., 2012; Vihman et al., 2006) corroborate the existence of cross-linguistic and developmental differences in the acquisition of interval duration. Grabe et al. (1999) examined the variability in vocalic duration in the speech of four-year-old French and English children and found that only the French children had acquired adult-like rhythmic patterns, with appropriate relative duration of final syllables. DePaolis and collaborators (DePaolis et al., 2008) studied disyllabic productions of French, English, Welsh and Finnish infants at the onset of speech and found that infants learning French or English, languages with prominent final syllable lengthening, exaggerated this feature and produced second syllables that were too long and with a high degree of variability. On the other hand, children learning Finnish or Welsh, languages with moderate final syllable lengthening, showed much more stability in the relative duration of the first and second syllable.

Payne, Post, Astruc, Prieto, and Vanrell (2010, 2012), using the same child participants as the present study concluded that from a very early stage in child speech, language-specific characteristics co-exist with properties common to child speech cross-linguistically, such as lower vocalic interval variability but higher consonantal variability than adult speech. Such characteristics are possibly the result of an incomplete mastery of prosodic timing on the one hand, and of consonantal articulation on the other. As these skills are mastered, cross-language similarities disappear, so that by six years the rhythmic characteristics are more closely aligned with those of adult speakers for each language.

Another source of variability that can influence the development of intonation arises from syllabic structure. It has been shown that differences in the statistical frequency of syllabic structures can have an influence on the acquisition of early phonological templates. Children learning Spanish, a language with relatively simple syllabic structures and a higher statistical frequency of words with three or more syllables, can produce multisyllabic words much earlier than children learning Catalan (Prieto, 2006), English or German (Lleó, 2006; Lleó & Demuth, 1999) as they do not have to contend with complex syllabic structures and they can concentrate on increasing the number of syllables. Similar observations have been made for other languages (e.g. Vigário, Freitas, & Frota, 2006, for Portuguese; Vihman, DePaolis, & Davis, 1998, for French and English).

Since alignment strongly depends on the segmental composition and duration of the stressed syllables, we expect that Spanish and perhaps also Catalan children will acquire the rhythmic pattern of their languages earlier than English children and that they may master adult-like alignment earlier as well. On the other hand, on the basis of reports in the literature (Allen & Hawkins, 1978, 1980; Kehoe et al., 1995; Prieto, Estrella, et al., 2012; Snow, 1994, 1995, 1998; Vanrell, Prieto, Astruc, Payne, & Post, 2011), we may expect pitch scaling to be quite stable from an earlier age in all three languages.

3 Methodology

We compared the intonational patterns of words controlled for number of syllables and stress position in the child speech to the adult target speech provided by their mothers, both elicited with a
controlled naming task. Recordings were made in Cambridge, Madrid and Barcelona, using professional quality recorders and microphones in the participants’ homes.

3.1 Participants and procedure

We recorded 36 children interacting with their mothers. All recordings were conducted at the children’s homes in sessions of about 40 minutes. There were 12 children per language (12 English, 9 girls and 3 boys; 12 Catalan, 7 girls and 5 boys; 12 Spanish, 7 girls and 5 boys) aged about two, four and six years of age at the time of the recordings. The ages of the children were chosen so as to fall into clearly differentiated developmental stages.2

Not all the children produced sufficient speech and some recordings contained background noise or presented other problems. In order to have comparably complete datasets for every language and age, we selected the best three recordings for every age range and language. We thus analysed for the present study 27 children and 12 adult women who acted as controls (see Table 1).

The data were elicited with a naming game, based on animated pictures shown on PowerPoint slides on a laptop screen. The animations showed scenes, with animals or everyday objects, that included the target word. Mothers were given written instructions asking them to read a short story about a fairy called Melanie who was looking for some objects and animals. According to the instructions, the mother asked her child to name the target words asking, What is Melanie looking for? or What is this? and then praised the child for getting it right, and repeated what the child had said. If the child said a different word, as for instance ball instead of the target word balloon, the mother was instructed to encourage her to try again until the child used the target word. The dialogue was modelled for the mother in each slide, with the target word highlighted in a different colour. A typical dialogue went thus:

(1)
[mother]  What is Melanie looking for?
[child]  The balloon
[mother]  Good! She is looking for the balloon.

Table 1. Participants’ characteristics and number of tokens in each language and age group.

<table>
<thead>
<tr>
<th>Language</th>
<th>Age</th>
<th>Age range</th>
<th>Gender</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>2</td>
<td>2:2–2:8</td>
<td>F,F,M</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3:9–4:11</td>
<td>F,F,M</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6:4–6:10</td>
<td>M,M,F</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Adults</td>
<td>25–40</td>
<td>F</td>
<td>23</td>
</tr>
<tr>
<td>Spanish</td>
<td>2</td>
<td>2:3–2:7</td>
<td>F,F,M</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3:5–4:5</td>
<td>M,M,F</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5:9–6:7</td>
<td>F,F,M</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Adults</td>
<td>25–40</td>
<td>F</td>
<td>61</td>
</tr>
<tr>
<td>English</td>
<td>2</td>
<td>1:11–2:0</td>
<td>F,F,M</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3:10–4:10</td>
<td>F,FF</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>5:10–6:0</td>
<td>F,FF</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Adults</td>
<td>25–40</td>
<td>F</td>
<td>44</td>
</tr>
</tbody>
</table>
At a later stage, we recorded an additional dataset of adult-directed speech with three adult female speakers of each language in the role of the mother, interacting with a native-speaker researcher.

### 3.2 Experimental material

We controlled the prosodic composition of the target words, using words with eight different stress patterns: S, SW, WS, WSW, SWW, WWS, SWSW, SWSWW. The words were selected so as to be imageable and familiar to young children. We used words such as *train, tren* and *tren* (S); *balloon, camión, camió* (WS); *baby, mono, mono* (SW), etc. (see word list in the Appendix). We also included in the analyses five pairs of monosyllabic and disyllabic words (e.g. *key* and *monkey*), originally recorded for a different experiment. While this means that there is a larger number of monosyllables and disyllables, this is in line with their higher frequency of occurrence in child speech in all three languages under investigation.

For English, we consulted the MRC Psycholinguistics Database (Wilson, 1988) to select the most familiar words for each of the intended prosodic patterns for the younger children and we filled the gaps by looking at the literature on L1 acquisition and at children’s books. For Spanish and Catalan there was no equivalent database available, so we translated the English words selected and filled the gaps using children’s books. The target words were selected on account of being easy to pronounce and as segmentally comparable as possible across languages; open and closed syllables were balanced across languages.

The age of the children posed some methodological problems. Because of the tendency for young children to simplify complex codas (Gerken, 1994; Vihman et al., 1998), we mostly used open syllables. Because of the possible effects of stressed syllable type in alignment (e.g. Prieto & Torreira, 2007; Schepman, Lickley, & Ladd, 2006), we also had to maintain a balance of open and closed syllables in the three languages across the whole database. There were between 7 and 10 target words with closed stressed syllables per language as follows (see phonetic transcriptions in the Appendix):

**English:** 9 words (sun, train, moon, angel, monkey, giraffe, balloon, violin, helicopter)

**Spanish:** 10 words (sol, tren, bus, camión, león, caracol, pantalón, violin, elefante, helicóptero)

**Catalan:** 7 words (sol, tren, bus, taxi [tak-si], cocodrilo, elefant, helicópter)

### 3.3 Intonational contour

For the intonational analysis of the data, we followed the labelling conventions of ToBI (Tones and Breaks Indices, see Silverman et al., 1992), the transcription system that complements the AM framework, and that has different versions for different languages. For Spanish, we used Sp_ToBI, the model proposed by Estebas-Vilaplana and Prieto (2008) and Estebas-Vilaplana and Prieto (2010). For Catalan, we used Cat_ToBI (Prieto, Aguilar Cuevas, Mascaro i Pons, Torres-Tamarit, & Vanrell, 2009). For English, we based our analyses on Gussenhoven’s proposal for Dutch (Gussenhoven, 2005), given that the intonational phonology of Dutch is very similar to that of English (see Gussenhoven, 2004, p. 297, for a description of the intonational phonology of English). One significant departure from the original ToBI model proposed by Gussenhoven (2004, 2005) that we have adopted in our analysis is the lack of phrase accents (L- or H-) associated to the end of phrases. Thus, our phrases end with just IP boundary tones.3
We chose for investigation an intonation contour that is comparable across the three languages and that occurred frequently in our database. This contour is characterised by an optional rise (onglide) from the beginning of the utterance to the high target, followed by a fall, as shown in Figure 1.

Figure 1. Schematic representation of two phonetic variants of the target contour in all three languages: (a) Type 1; (b) Type 2.

In Spanish and Catalan the initial rising onglide (represented by the L+H* accent) is generally present, whereas in English it seems to be optional, in free variation, as maintained by Grabe (1998, Chapter 4). For instance, the rising onglide may be absent in short words with voiceless material on the stressed syllable like tea, or cockatoo, which do not provide enough segmental material to realise the initial rising movement. However, the rising onglide can be omitted even in contexts where there is enough voiced material. The difficulty in distinguishing between L+H* and H* has been noted repeatedly in the literature on English intonation. Pitrelli, Beckman, and Hirschberg (1994) carried out an evaluation of inter-transcriber agreement in English ToBI and found that the distinction between these two categories was a frequent point of disagreement. They interpreted L+H* as a minor variant of H* and decided to merge the two categories in their analysis of the data. Ladd (1996, p. 84) comments that ‘[o]n a phrase initial accented syllable, L+H* and H* can be difficult to distinguish’. Gussenhoven (2004, p. 298) analyses the L as an optional low onset that aligns to the start of the accented syllable.

In our data we found another pattern, one in which the H* tone is sustained and extended onto a plateau, often with a very level pitch range, as represented in Figure 2.

Figure 2. Schematic representation of L+H* with a plateau (Type 3 accent) in all three languages.

This configuration is extremely frequent in the child data, especially for the younger children, although the mothers also use it occasionally. The appropriate phonological analysis of the plateau contour is still unresolved. For instance, Kent and Murray (1982) report 31% of contours with a plateau (‘level tones’) in their data, and they discussed whether to group them with the falls or not. Snow and Balog (2002) discuss whether level tones are a remnant of pre-linguistic intonation. Our hypothesis is that children attempted the L+H* pitch accent, which is the appropriate intonational choice in the context, but they undershoot the rising and falling movements. Thus, we may
consider (L)+H* as the underlying pitch accent in the plateau contour, although further testing would be needed to validate this hypothesis.

4 Analysis

The recordings were digitised and processed using Praat by English-, Catalan- and Spanish-speaking research assistants. The intonational patterns of the utterances were analysed by the authors, by auditory analysis and inspection of the F0 traces.

All vowels and consonants in the target word were segmented and labelled as either C (consonant) or V (vowel), following these segmentation procedures: the beginning or end of a sonorant consonant was identified at the start of the abrupt change from the steady-state period in the spectrogram to the onglide transition movement to the vowel. When the formant transitions were not abrupt enough, the criterion used was the change in amplitude displayed in the waveform. Voiceless stops were segmented from the burst in all cases, for comparability with cases in which they were word-initial.

In the database we coded the loudness level of the speaker for each token (1 quiet or shy, 2 normal, 3 loud or emphatic). We also coded the prosodic pattern of each word (S, SW, WS, etc.) as it was actually realised by the child, since truncation is relatively frequent in early child speech (see an analysis of truncation in this same database in Astruc, Payne, Post, Prieto, & Vanrell, 2010). We also labelled the indefinite article in Catalan and Spanish (un, una) and we coded it in the database, since it is known that it is stressed in Catalan and Spanish (Hualde, 2005, p. 234) and that it often receives a pitch accent that can trigger downstep (lowering in scaling) of the following pitch accents.

Two segmental and five tonal targets were marked by hand by the authors, using the following segmental and tonal labels:

(2) Segmental labels
O  Onset of the accented syllable
E  End of the accented syllable

(3) Tonal labels
H*  Peak location in Type 1 and Type 2 accents
H1  Peak location in Type 3 accents
H2  End of the plateau in Type 3 accents
L  F0 valley location in all accents
L%  Lowest measureable F0 point at the end of the utterance

The tonal and segmental targets were located visually and acoustically using a display of waveforms, wide-band spectrograms and F0 tracks obtained with Praat 4.2 (Boersma & Weenink, 2004) and were annotated on a text file.

In the case of Type 3 accents where the high targets formed a plateau with no clear F0 maximum, following standard procedures (Ladd et al., 2000) H1 was placed at the highest F0 around the end of the accented syllable of the test word (as was H* in the other two accent types). Following research on perception of tonal plateaux by Knight and Nolan (2006), H2 was marked at the point where pitch starts to drop (the falling elbow), as this is the point estimated to be perceptually relevant for hearers. Micro-prosodic effects (such as the dip produced by nasal segments) were disregarded.

An example of the measurement points that illustrates the labelling scheme used in the three accent types described above, and represented in Figures 1 and 2, is shown in Figure 3. All three
panels correspond to the target word money, uttered by three English children, two of them aged four and one aged two (who produce the word as ‘monies’ [ˈmʌnɪs]).

Pitch scaling (pitch range) was measured as the difference between the F0 in semitones at H* and the F0 in semitones at L%, the lowest value at or near the end of the utterance. The target words were nuclear accents in the phrase-final position and were obtained with a task that elicited identificational narrow focus on the nuclear syllable. From observations and descriptions in the literature, in Spanish and in Catalan we expected both the L and the H* targets to be aligned with the stressed syllable (Estebas-Vilaplana & Prieto, 2010). In nuclear accents, the L is anchored at the syllable onset and the H* towards the end of the stressed syllable (in prenuclear accents the H* is aligned with the post-tonic syllable; Estebas-Vilaplana, 2009, for Catalan; Prieto, van Santen, & Hirschberg, 1995, for Spanish). We may expect that in syllables with sonorant codas, the H* may be aligned earlier in the syllable (Prieto & Torreira, 2007). In English, the H* peak is aligned very close to the beginning of the stressed vowel and is variably situated between 19 ms after the vowel onset and 97 ms before the vowel offset (Ladd et al., 2009). The L, when present, is aligned with the onset of the stressed syllable (Gussenhoven, 2004, p. 297).

**Figure 3.** Pitch traces for English [ˈmʌni] and [ˈmʌniz] with (a) Type 1 and (b) Type 2 accents. (c) Pitch trace of [ˈmʌni] with Type 3 accent.
Three different measures of L and H alignment were used: (i) L to Syllable Onset; (ii) H* to Syllable Onset; (iii) H* to Syllable End. All three measures have been used in the literature on alignment. (i) is appropriate for the data because the leading L tone is anchored at the onset of the stressed syllable. For instance, Prieto et al. (1995) used both (ii) H* to Syllable Onset and (iii) H* to Syllable End. Xu (1998) and Xu and Liu (2006) proposed that syllabic onsets provide a more robust anchoring point for tonal movements, thus providing arguments for (ii). Schepman et al. (2006) compared several measures of alignment, among them (ii) and (iii), and concluded with a recommendation to use ‘local’ measures whenever possible (see also Atterer & Ladd, 2004; Prieto & Torreira, 2007). In the present data, we expect great variability in the duration and structure of the syllables, which may trigger variations in the alignment of the H*. For this reason we decided to use both measures of H* alignment, H* to Syllable Onset and H* to Syllable End, in addition to L to Syllable Onset.

Finally, in order to account for stressed syllable type effects on the alignment of the tonal targets, we listened to each token and annotated the actual realisation of the stressed syllable using these labels:

(4) Stressed syllable type labels
(P)V(P) No sonorant onset or coda
SV Sonorant onset
(C) VS Sonorant coda
SVS Both sonorant onset and coda

5 Results

The original database contained 624 tokens. Prior to the analyses, we explored these data using box-plots and outliers and extreme values were identified and re-measured by hand, but 27 tokens had to be discarded as they were shouted or had F0 perturbations.

The final data set had 597 tokens, of which 368 tokens were produced with a falling tone, with or without an initial L onglide and 229 with an extended plateau, the Type 3 tone of Figures 2 and 3. The plateaux were mostly produced by the younger children, as we see from Figure 4.

Two-year-olds produced 39% of the plateaux, followed by four-year-olds with 31%, and then by six-year-olds. Adults, on the other hand, produced only 8% of plateaux. These results are in line with previous findings on the prevalence of level tones in early child speech (e.g. Kent & Murray, 1982), although the pitch range of the plateaux in our data is not always completely level.

Figure 4. Pie chart showing the distribution of Type 3 accents (as a percentage of the total) by age group.
In the next section, we analyse quantitatively the scaling and then the alignment of the remaining tokens that were produced with the falling tone.

5.1 Scaling

We measured scaling in the child dataset (258 cases) from which we excluded 39 instances of shouting or very faint speech. The child dataset thus had 219 cases (age 2 = 60, age 4 = 61, age 6 = 98). In order to compare the child productions to ‘normal’ adult speech, a new dataset of adult-directed speech (ADS) was collected. For each language three adult women were recorded performing the same tasks together or with the researchers, making a total of 242 cases, which, together with the child dataset of 219 cases, made a total of 461 cases used in the analyses of scaling.

We have calculated the scaling of the accent as the pitch range difference between the H* target and the L%. Figure 5 shows the pitch range for each language and age group.

![Figure 5](image)

*Figure 5.* For each age group from left to right, ages 2, 4, 6, and adults (on the horizontal axis) the panel shows the F0 difference (in semitones) between H* and L% (on the vertical axis). The bars show the mean and the standard error of the mean for English, Spanish and Catalan.

On average, the data of the two-year-olds have a pitch range of 7 st, four-year-olds, 6.2 st, six-year-olds, 4.3 st. Adults using ADS have a pitch range of 5 st. An analysis of variance (ANOVA) with dependent variable pitch range and between-subjects factors Age and Language confirmed statistically significant effects of Age, $F(3, 455) = 16.581$, $p = 0.001$, partial eta-squared = 0.099; Language, $F(2, 455) = 9.594$, $p < 0.001$, partial eta-squared = 0.040; and an Age by Language interaction, $F(6, 455) = 4.909$, $p < 0.001$, partial eta-squared = 0.061. The younger children (two- and four-year-olds) have a wider pitch range than the older children and adults (the differences in pitch range between six-year-olds and adults are not statistically significant).

In order to check whether the position of the stress had an effect on the pitch range of the younger children, we broke up the data according to Age, Language and Stress (oxytones and non-
Oxytone words have the stress on the last syllable or are monosyllables (e.g. S, sun, WS guitar, WWS kangaroo). Non-oxytone words include all the paroxytones (SW, baby) and proparoxytones (SWW, butterfly) in the database.

It has been reported in the literature (e.g. Allen & Hawkins 1978, 1980; Snow 1995, 1998), that young children have yet to learn how to control pitch properly, while adults and more linguistically mature children show a robust scaling of the pitch accent that is quite constant across different stress patterns. Adults can achieve the appropriate scaling even in word- and phrase-final position, when there is less space for the tonal movement to unfold and there is a risk of not reaching the target. Allen and Hawkins (1978) analysed monosyllabic and disyllabic words from three children at age three and concluded that the children produced the stressed syllables of S and WS words with a more reduced pitch range than those of SW words. Snow (1995, 1998), on the other hand, found that 18–24-month-old English children were able to produce a similar accent range for the nuclear (L)+H* accent of SW and S words (e.g. money and knee). Figure 6 shows the pitch range for each age group and for each stress pattern.

![Figure 6](image.png)

**Figure 6.** Mean F0 difference in semitones between H* and L% (on the vertical axis) for oxytones and non-oxytones (on the horizontal axis). The bars show the mean and the standard error of the mean for each age group from left to right, ages 2, 4, 6, and adults.

An ANOVA with dependent variable pitch range and between-subject factors Stress, Language and Age confirmed statistically significant effects of Age, $F(3, 443) = 17.646, p < 0.001$, partial eta-squared = 0.107; Language, $F(2, 443) = 11.463, p < 0.001$, partial eta-squared = 0.049; and an Age by Language interaction, $F(6, 443) = 5.278, p < 0.001$, partial eta-squared = 0.009. Stress, on the other hand, is not significant and there are no significant interactions with Age or Language. This means that the difference in pitch range between oxytones and non-oxytones is not statistically significant overall or for any age group or language; even the younger children can control their pitch range to compensate for the proximity to the end of the word and the utterance in oxytones, which corroborates the findings of Snow (1995, 1998).
In summary, two- and four-year-olds have a wider pitch range than six-year-olds and adults. However, they have already learnt how to compensate for the fall in pitch that occurs at the end of an utterance.

5.2 Alignment

We analysed the alignment of the H* peak in the 368 tokens produced with Type 1 and Type 2 accents (see Figures 1 and 3). We did not include in the analysis of the alignment utterances with two pitch accents as this would be likely to affect the alignment of both. We used three measures of L and H alignment: (i) L to Syllable Onset; (ii) H* to Syllable Onset; (iii) H* to Syllable End.

The error bars in Figure 7 show the alignment of the L to the syllable onset for Spanish and Catalan. The alignment of the L target could be computed only for Spanish and Catalan, since most of the English data lacks the onglide rising movement (as described in Section 3) and hence also lacks a measurable L target. As we see in the error bars in Figure 7, even the youngest children show a remarkably precise alignment of the L target.

![Figure 7](image)

**Figure 7.** Distance (in seconds) from the L target to the syllable onset (on the vertical axis) for each age group (on the horizontal axis). The bars show the mean and the standard error of the mean for Spanish and Catalan.

In order to further explore the data, we ran a series of analyses of covariance (ANCOVAs) for each alignment measure with between-subjects factors Age and Language and covariate factor Syllable Duration. The analyses confirm statistically significant effects of Syllable Duration, $F(1, 261) = 16.504, p < 0.001$, partial eta-squared = 0.061; Age, $F(3, 262) = 3.726, p = 0.012$, partial eta-squared = 0.042, but not of Language, although the interaction of Age by Language is significant, $F(3, 262) = 5.080, p = 0.002$, partial eta-squared = 0.057. Catalan two-year-olds show a slightly less accurate alignment than the Spanish children, but this difference is extremely small (less than 5 ms from the adult target). The older Catalan children show differences with the adult target of 4 ms or less, which are similar to those between the Spanish age groups and are not statistically significant.
As for the alignment of the H* peak, we see in the error bars in Figure 8 and 9 that adults are much closer to the target (the syllable onset and end, respectively) than children are. In both figures, we see as well cross-linguistic differences in H*.

In both figures, we see as well cross-linguistic differences in H to Syllable Onset and H to Syllable End. As we see in Figure 9, Catalan and Spanish adults align their H* target slightly closer to
the end of the syllable than English speakers do. We also see that Spanish two-year-olds show less variability in their alignment and are much closer to the adult target than are Catalan and English two-year-olds. As Catalan and English children grow older, they align their H* increasingly closer to the adult target and, eventually, Catalan six-year-olds catch up with the Spanish children. (We have to bear in mind, however, that the English children were younger than the Catalan and Spanish children, and this may have an important effect – see Note 1.) We notice as well much more variability, especially for the younger children, in the alignment of H* to Syll Onset (in Figure 8) than on the alignment of H* to Syll End (in Figure 9). We hypothesise that children lengthen their stressed syllables much more than adults do and this in turn affects alignment.

To examine the effects of syllabic duration on the alignment of the H* target, we ran a series of ANCOVAs for each H* alignment measure with between-subjects factors of Age and Language and with a covariate factor of Syllable Duration. For the alignment of H* to Syllable Onset, the analyses confirm statistically significant effects of Syllable Duration, $F(1, 355) = 68.486, p < 0.001$, partial eta-squared = 0.162; Age, $F(3, 355) = 10.642, p < 0.001$, partial eta-squared = 0.083; and Language, $F(2, 355) = 5.525, p = 0.004$, partial eta-squared = 0.030. The interaction between Age and Language almost reaches statistical significance.

As for the alignment of H* to Syllable End, there are statistically significant strong effects of Syllable Duration, $F(1, 355) = 864.775, p < 0.001$, partial eta-squared = 0.709, and weaker effects of Age, $F(3, 355) = 8.488, p < 0.001$, partial eta-squared = 0.067; Language, $F(2, 355) = 4.172, p = 0.016$, partial eta-squared = 0.023; and of the Age by Language interaction, $F(6, 355) = 2.503, p = 0.022$, partial eta-squared = 0.041. Comparing the effect sizes obtained for both measures of peak alignment, we see that Duration is the only variable that has any strong effects on alignment (partial eta-squared = 0.709) and this is in the alignment of H* to Syllable End. The alignment of H* to Syllable Onset is affected much less by the duration of the syllable (partial eta-squared = 0.162). Therefore, at least in this type of data, H* to Syllable Onset is the more robust measure of H* peak alignment.

In the following round of analyses, we break up the data according to the stress pattern of the target words (oxytone, paroxytone, proparoxytone). Figures 10 and 11 show the alignment of H* to Syll Onset and to Syll End (respectively) for each age group and for each stress pattern in relation to syllable duration.

As we see in Figures 10 and 11, the duration of the stressed syllable has an effect on the alignment of the high tonal targets. There is a significant correlation between both measures of alignment and the duration of the syllable, but this is much stronger in H* to Syllable End (Pearson = −0.812, $p < 0.001$) than in H* to Syllable Onset (Pearson = −0.556, $p < 0.001$). Children produced stressed syllables with a mean duration of 0.395 s at age two, 0.406 s at age four, 0.305 s at age six. Adults produced syllables with a mean duration of 0.228 s. We tested this with an ANOVA with the dependent variable Syllable Duration and the factors Age and Language for all the words in the dataset. The results confirmed statistically significant effects of both Age and Language: Age, $F(4, 609) = 185.638, p < 0.001$, partial eta-squared = 0.149; Language, $F(2, 609) = 5.197, p = 0.006$, partial eta-squared = 0.031; and also an Age by Language interaction, $F(5, 609) = 3.947, p = 0.002$, partial eta-squared = 0.031. In English, the syllables of the younger children (age two, 0.407 s; age four, 0.340 s) are much longer than those of the older children (0.301 s) and the adults (0.234 s). In both Spanish and Catalan, the differences are smaller and in Spanish they are statistically significant only for the two-year-olds and the adults.

The stress pattern of the word has an effect on the alignment of the H* peak in both measures: in oxytone words the peak is retracted, while in paroxytones and proparoxytones the peak is aligned near the end of the syllable (although the exact location depends strongly on syllabic duration). Adults and six-year-old children show less difference and less variability in the
Figure 10. Distance (in seconds) from H* to the syllable onset (on the vertical axis) and the duration of the syllable (in seconds, on the horizontal axis) for each age group. (a) and (b) correspond to ages two and four; (c) and (d) to ages six and adults. The target words are separated in oxytones (a circle), paroxytones (a triangle) and proparoxytones (a star).

Figure 11. Distance (in seconds) from H* to the syllable end (on the vertical axis) and the duration of the syllable (in seconds, on the horizontal axis) for each age group. (a) and (b) correspond to ages two and four respectively; (c) and (d) to ages six and adults. The target words are separated in oxytones (a circle), paroxytones (a triangle) and proparoxytones (a star).
alignment of oxytones, paroxytones and proparoxytones than younger children do. For the two- and four-year-olds, the difference between oxytones and between and paroxytones and proparoxytones is of the order of 100 ms; for adults and six-year-olds it is about 50 ms.

These observations are confirmed by two ANCOVAs with dependent variables H* to Syll Onset and H* to Syll End, with between-subjects factors Age and Stress and covariate factor Syllable Duration. For the alignment of H* to Syll Onset, the analyses confirm statistically significant effects of Syllable Duration, $F(1, 313) = 122.288, p < 0.001$, partial eta-squared = 0.281; Stress, $F(2, 313) = 35.360, p < 0.001$, partial eta-squared = 0.184; and Age, $F(3, 313) = 4.457, p = 0.004$, partial eta-squared = 0.041. The interaction of Age by Stress is also significant, $F(6, 313) = 2.389, p = 0.028$, partial eta-squared = 0.044. As for the alignment of H* to Syll End, we find significant effects of Syllable Duration, $F(1, 313) = 449.004, p < 0.001$, partial eta-squared = 0.589; Stress $F(2, 313) = 45.441, p < 0.001$, partial eta-squared = 0.225; Age, $F(3, 313) = 2.460, p < 0.001$, partial eta-squared = 0.023; and of the Age by Stress interaction, $F(1, 313) = 2.851, p = 0.010$, partial eta-squared = 0.052.

Comparing the effect sizes, we see that Stress has a large effect on the alignment of H* to Syll End (partial eta-squared = 0.225) and a medium one on the alignment of H* to Syll Onset (partial eta-squared = 0.184), and that this effect is larger in older children and in adults. Effectively, the accented syllable of oxytone words is the last one in the word and, in this data, also the last in the phrase, and is thus substantially longer than other syllables. It is remarkable that even the younger children align their targets correctly near the end of the stressed syllable; they thus have already acquired adult-like patterns of alignment. Their problem seems to be that they produce stressed syllables that are still too long and this excessive syllable duration may in turn influence the alignment of the H* target.

In order to rule out potential effects of syllable duration, syllable structure and stress, we conducted a multivariate ANOVA with a covariate factor (MANCOVA) across the three languages and the four age groups with the three alignment measures (L to Syll Onset, H* to Syllable Onset, H* to Syllable End) as dependent variables and duration as a covariate factor. The within-items factors were Syllable Structure (four levels: no sonorant onset or coda, sonorant onset, sonorant coda and both sonorant onset and coda) and Stress (three levels: oxytones, paroxytones and proparoxytones). The results confirm that duration has significant effects on the two measures of alignment of the H target (in decreasing order of strength): H* to Syll End, $F(1, 299) = 348.465, p < 0.001$, partial eta-squared = 0.567 and H* to Syll Onset, $F(1, 299) = 155.225, p < 0.001$, partial eta-squared = 0.369. For the H* target, the longer the syllable, the less precise the alignment with the segmental target. This effect is rather strong, especially when alignment is measured in relation to the end of the syllable. For the L target, the effect is statistically significant but small: L to Syll Onset, $F(1, 299) = 7.752, p = 0.006$, partial eta-squared = 0.028. Stress has significant but modest effects on the alignment of the H* target, H* to Syll Onset, $F(2, 299) = 23.079, p < 0.001$, partial eta-squared = 0.139 and H* to Syll End $F(2, 299) = 28.677, p < 0.001$, partial eta-squared = 0.167), but not on the alignment of the L target. On the other hand, the type of syllable structure has no significant effect on the three alignment measures, even in the adult data. If the presence of a sonorant onset and/or coda, and/or the stress pattern of the word have any effects on alignment, these may be realised via the duration of the syllable.

In summary, we found statistically significant effects of age, language and duration. Comparing the alignment of oxytones, paroxytones and proparoxytones we found that younger children show similar patterns of alignment to adults and older children; in oxytones the peak is retracted, while in paroxytones and proparoxytones the peak is aligned near the end of the syllable. The Catalan and English children show less precision in their alignment in comparison to adults than the Spanish
children do. However, the factor that has the strongest effects on alignment is the duration of the stressed syllable.

6 Discussion and conclusion

This study set out to investigate whether young children can control the phonetic aspects of intonation in terms of pitch scaling and pitch alignment and to identify any cross-linguistic differences. Our methods and findings are in line with recent theories of phonological acquisition and child phonology. We assumed that children are sensitive to the syllable and perhaps to the more coarse-grained syllabic components (e.g. Ziegler & Goswami, 2005), to the metrical template of the word (Vihman & Croft, 2007) and to the prosodic context around the tonal target (Pierrehumbert, 2003), and we designed our test material accordingly. We controlled for stress position and, as far as was possible, for syllable type (favouring open syllables as much as possible).

Performing a fine-grained quantitative analysis of the scaling and alignment of tonal targets based on the AM method, we found that the younger children can control tonal alignment and scaling from a very young age, although they do not reach adult-like proficiency until later. As for scaling, and in line with previous studies, we found that younger children did produce a very large proportion of plateaux with a fairly level pitch range. Older children use these level contours less and less, and they are virtually non-existent in the adult speech, at least in styles other than child-directed speech.

We found cross-linguistic as well as age-related differences in the pitch range of the words produced with a (L)+H* nuclear accent. We observed that English children use a wider pitch range than do Catalan and Spanish children of the same age. We have also noticed that six-year-olds have a pitch range similar to that of adults; six-year-olds are thus closer to ADS. In line with previous findings in English (Snow, 1995, 1998), we found that whether the word is an oxytone or not has no significant effects on pitch range. Children control pitch well enough to produce a similar pitch range for paroxytones (money) and oxytones (knee), just as adults do.

We compared the alignment of the nuclear peaks in child speech to that of the adult data, using two measures of H* alignment (H* to Syll Onset and H* to Syll End) and one measure of L* alignment (L* to Syll Onset). Both H* measures have been used in the literature and it is still undecided which one of them is the more robust. Our results show that H* to Syll Onset is preferable for measuring nuclear peaks in data similar to ours; child data and adult data are both affected by variations in duration. As shown by the effect sizes, the duration of the stressed syllable has a strong effect on alignment. This effect is smaller on the alignment H* to Syll Onset (partial eta-squared = 0.281) than on the alignment H* to Syll End (partial eta-squared = 0.589). For both measures the stress pattern of the word has a statistically significant effect that is smaller for H* to Syll Onset (partial eta-squared = 0.184) than for H* to Syll End (partial eta-squared = 0.225), which is a large effect. In conclusion, H* to Syll Onset is the more robust measure for the alignment of nuclear accents in the present English, Spanish and Catalan data. This is in line with the recent findings for two varieties of English reported by Ladd et al. (2009).

Our findings offer further corroboration that young children have difficulty controlling the timing aspects of prosody, as has been observed in the literature (from Allen & Hawkins, 1978, to Payne et al., 2012). Indeed, the children in the present study were also analysed doing a different task by Payne et al. (2010, 2012), who report excessive lengthening of final syllables. In our data, we find that, overall, the younger children produce stressed syllables that are much longer than those of the adults, but the differences are much bigger for English, a language
with vocalic reduction of unstressed syllables and prominent final lengthening effects. In all three languages there are large statistically significant differences between adults and children at two and at four years old. In English, however, differences are larger (almost double than the adults, in the case of the younger children), and are statistically significant for all age groups.

Despite their lack of mastery of syllable durations, even the youngest (Catalan and Spanish) children produced adult-like alignment of the L target right from the start. Although their alignment of the H* target is not as accurate as that of adults, this is because of the effects of syllable duration. In fact, if we examine the target words according to stress pattern (oxytones, paroxytones and proparoxytones) we see that the two-year-olds align the H* target to the end of the syllable very accurately in proparoxytones and paroxytones, much more than they do in oxytones. In oxytones, they have to fit two tonal targets (the H* and the L% boundary) into just one syllable, and as a consequence of this tonal crowding, the H* target is realised earlier in the syllable. However, the fact that at such a young age children have acquired the basic alignment constraint, that is, the constraint to realise both tonal targets in one syllable, is in itself remarkable.

Our results fit well with recent proposals that claim that prosodic encoding in child phonology is learned together with segmental encoding (Pierrehumbert, 2003; Vihman & Croft, 2007). Taking a bottom-up statistical approach to language acquisition, Pierrehumbert (2003) proposes that children acquire the phonology by first storing the input in all its rich phonetic and prosodic detail. Vihman and Croft's (2007) model agrees with Pierrehumbert's in claiming that learning phonology implies acquiring language-specific phonotactic templates (e.g. CV-CV). Astruc et al. (2010) analysed truncation patterns in part of the same dataset discussed here and found that two- and four-year-old children truncated syllables and omitted or swapped segments: they were still far from achieving adult-like segmental phonology. The present data show that the same young children are remarkably good at aligning tonal targets with the onset and offset of syllables, even though they cannot yet control syllable duration. We interpret these results as an indication that children acquire templates with coarse-grained phonotactic (e.g. including syllabic onsets and offsets, but not segments) and prosodic information (including stress, intonation and some aspects of rhythm), which are later filled in with finer-grained segmental information.

Finally, we have some support for our initial assumption that the prosodic typology of the L1 language would influence the acquisition of intonation. Both in the analysis of pitch range and of the alignment of the H* target, we found small but significant Age by Language interactions. Young Spanish children show a more precise command of the alignment of the high targets than Catalan and English children, but it is also true that the English children were younger and this may have had an effect, especially at age two, which is a period of rapid phonological development. We thus have to interpret these apparent cross-linguistic differences with caution. Both in alignment and in pitch range the important fact is that children as young as two are remarkably close to the adult model in the language they are learning: they can control the phonetic implementation of intonation in its dual aspect of pitch height and pitch timing, although only the oldest children can control syllable duration adequately.

**Acknowledgements**

We are grateful to Núria Argemí, Anna Barberà, Melanie Bell, Ana Estrella and Francesc Torres-Tamarit for recording the data in different languages, and especially to Naomi Hilton for segmenting the data. Some preliminary results of this study were presented at Tone and Intonation in Europe 2008 and have been published.
as university working papers (Astruc, Payne, Post, Prieto, & Vanrell, 2009). This article has been greatly improved thanks to the work of the Associate Editor, Marilyn Vihman, and our two reviewers, Sam Hellmuth and Carol Stoel-Gammon. Any errors that remain are ours.

**Funding**

This work was supported by a Batista i Roca grant (Ref: 2007 PBR 29, P.I. Pilar Prieto); by a British Academy Small Research Grant (Ref: SG-5177; P.I. Brechtje Post); by a research grant awarded by the Spanish Ministerio de Ciencia e Innovación (FFI2009-07648/FILO); and by a grant awarded by the Generalitat de Catalunya to the Grup d’Estudis de Prosòdia (2009SGR-701).

**Notes**

1. We use the terms ‘syllable-timed’ and ‘stress-timed’ merely as convenient shorthand to denote two clusters of rhythmic percept that have emerged from discrimination studies, while acknowledging that the factors leading to these percepts are far more complex and certainly not actually due to syllable-based or foot-based isochronicity.

2. As noted by the reviewers, the English children were younger overall than the Catalan and Spanish children. This may cause more difficulties for comparing data at age two, an age of rapid development, than at ages four and six, and means that some caution is required when interpreting any possible cross-linguistic differences at this age.

3. A notational discrepancy between Gussenhoven's AM model for English and the models for Spanish and Catalan is that the former only includes bitonal accents with trailing tones (T*+T), while the latter includes bitonal accents with both leading and trailing tones (T+T* and T*+T). This discrepancy has no effect on our analysis, however.

**References**


### Appendix

<table>
<thead>
<tr>
<th>Stress pattern</th>
<th>English</th>
<th>Spanish</th>
<th>Catalan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S</strong></td>
<td>Sun ['san]</td>
<td>Sol ['sol]</td>
<td>Sol ['sol]</td>
</tr>
<tr>
<td></td>
<td>Train ['trein]</td>
<td>Tren ['tren]</td>
<td>Tren ['tren]</td>
</tr>
<tr>
<td></td>
<td>Bee ['bi]</td>
<td>Pe ['pe]</td>
<td>Pe ['pe]</td>
</tr>
<tr>
<td></td>
<td>Moon ['mʊn]</td>
<td>Bus ['bus]</td>
<td>Bus ['bus]</td>
</tr>
<tr>
<td></td>
<td>Tea ['tɪ]</td>
<td>Sí ['si]</td>
<td>Sí ['si]</td>
</tr>
<tr>
<td></td>
<td>Knee ['nɪ]</td>
<td>Be ['be]</td>
<td>Be ['be]</td>
</tr>
<tr>
<td></td>
<td>Key ['kɪ]</td>
<td>Pe ['pe]</td>
<td>Pe ['pe]</td>
</tr>
<tr>
<td><strong>SW</strong></td>
<td>Angel ['eɪn.dʒ̩oʊl]</td>
<td>Nena ['ne.na]</td>
<td>Nena ['ne.na]</td>
</tr>
<tr>
<td></td>
<td>Baby ['beɪ.bi]</td>
<td>Luna ['lu.na]</td>
<td>Lluna ['lu.na]</td>
</tr>
<tr>
<td></td>
<td>Lion ['laɪ.ən]</td>
<td>Mono ['mo.no]</td>
<td>Mono ['mo.no]</td>
</tr>
<tr>
<td></td>
<td>Monkey ['mæ.ŋ.ki]</td>
<td>Llave ['xa.βe]</td>
<td>Taxi ['tak.si]</td>
</tr>
<tr>
<td></td>
<td>Potty ['pə.tɪ]</td>
<td>Papi ['pa.pi]</td>
<td>Papa ['pa.pa]</td>
</tr>
<tr>
<td></td>
<td>Money ['mæ.ni]</td>
<td>Bibe ['bi.βe]</td>
<td>Bibe ['bi.βe]</td>
</tr>
<tr>
<td><strong>WS</strong></td>
<td>Guitar ['ɡə.tər]</td>
<td>Bebé ['be.βe]</td>
<td>Bebé ['be.βe]</td>
</tr>
<tr>
<td></td>
<td>Balloon ['bɑ.ˈlʊn]</td>
<td>Camión ['ka.ˈmjon]</td>
<td>Camió ['ka.ˈmjo]</td>
</tr>
<tr>
<td></td>
<td>Giraffe ['ɡə.ˈræf]</td>
<td>León ['le.ˈon]</td>
<td>Lleó ['ʎə.ˈo]</td>
</tr>
<tr>
<td></td>
<td>Potato ['pɑ.ˈteɪ.tu]</td>
<td>Zapato ['扎实.ˈpa.to]</td>
<td>Sabata ['sa.ˈβa.to]</td>
</tr>
<tr>
<td></td>
<td>Banana ['bɑ.ˈnʌ.ˈna]</td>
<td>Pelota ['pe.ˈlo.ta]</td>
<td>Pilota ['pi.ˈlo.ta]</td>
</tr>
<tr>
<td></td>
<td>Pajamas ['pɑ.ˈdʒɑ.ˈmɑz]</td>
<td>Pijama ['pi.ˈxa.ma]</td>
<td>Pijama ['pi.ˈja.ma]</td>
</tr>
<tr>
<td><strong>SWW</strong></td>
<td>Elephant ['ə.ˈli.ʃənt]</td>
<td>Águila ['a.ˈɣi.la]</td>
<td>Aliga ['a.ˈli.ɣa]</td>
</tr>
<tr>
<td></td>
<td>Melanie ['mela.ˈni]</td>
<td>Mélanie ['me.la.ˈni]</td>
<td>Mélanie ['me.la.ˈni]</td>
</tr>
<tr>
<td></td>
<td>Butterfly ['bæ.ˈflai]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WWS</strong></td>
<td>Cockatoo ['kʊ.ˈko.ˈtʊ]</td>
<td>Caracol ['kα.ɾa.ˈkol]</td>
<td>Cocodrilo ['ku.ˈku.ɾo.ˈɾiɾ]</td>
</tr>
<tr>
<td></td>
<td>Kangaroo ['kæŋ.ˈɡɑ.ɾu]</td>
<td>Pantalón ['pan.tɑ.ˈlon]</td>
<td>Pantaló [ns] ['pan.tɑ.ˈlo(ns)]</td>
</tr>
<tr>
<td><strong>SWSW</strong></td>
<td>Caterpillar ['kæ.ˈpæ.tɪ.ˈpæl]</td>
<td>Mariposa ['ma.ɾi.ˈpo.sa]</td>
<td>Papallona ['pa.ˈpo.ˈka.ɾo]</td>
</tr>
<tr>
<td></td>
<td>Helicopter ['hel.ɪ.ˈkar.ət]</td>
<td>Cocodrilo ['ko.ko.ˈɾo.ɾi.lo]</td>
<td>Hipopótamo ['i.ɾu.ˈpa.ɾo]</td>
</tr>
<tr>
<td><strong>SWSWW</strong></td>
<td>Hippopotamus ['hi.ˈpɔ.ˈpə.tə.mɔs]</td>
<td>Hipopotamo</td>
<td>Hipopótamo ['i.ɾu.ˈpa.ɾo]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Helicóptero ['e.ˈli.ˈkɔp.ɾe.ro]</td>
<td></td>
</tr>
</tbody>
</table>

*In all cases we assume a CV syllabification, as this facilitates the comparison between English, Spanish and Catalan. An alternative syllabification words like *potty* would be ['pɒ.tɪ], with the /t/ grouped with the stressed syllable.*