Tonal targets in early child English, Spanish, and Catalan

Lluïsa Astruc^a, Pilar Prieto^b, Elinor Payne^c, Brechtje Post^d, Maria del Mar Vanrell^e

- a Department of Languages, The Open University and Department of Spanish and Portuguese, University of Cambridge
- b Institució Catalana de Recerca i Estudis Avançats and Departament de Filologia Catalana, Universitat Pompeu Fabra
- c Phonetics Laboratory, University of Oxford
- d Research Centre for English and Applied Linguistics, University of Cambridge
- e Departament de Filologia Catalana, Universitat Pompeu Fabra

Corresponding author

Lluïsa Astruc

Department of Languages

Faculty of Education and Language Studies

The Open University

Milton Keynes MK7 6AA

l.astruc@open.ac.uk

Running head: Tonal targets in early child Catalan, Spanish, and English

Acknowledgments

This research has been funded by a Batista i Roca grant (Ref: 2007 PBR 29, Pilar Prieto as P.I.) and by a British Academy Small Grant (Ref: SG-51777, P.I. Brechtje Post). 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.

Abstract

This study analyzes the scaling and alignment of low and high intonational targets in the speech of 27 children; nine English, nine Catalan, and nine Spanish, between the ages of two and six years. We have compared the intonational patterns of words controlled for number of syllables and stress position in the children's speech to the adult target provided by their mothers, both elicited with a controlled naming task. A total of 624 utterances were analyzed following the Autosegmental Metrical framework. Our results show that children as young as two can control relevant intonation parameters such as pitch height and pitch timing, although they still do not control syllabic duration. Even the youngest children show adult-like alignment of the low target, but their mastery of the high target increases with age. The prosodic typology of the ambient language appears to influence the acquisition of tonal targets; young Spanish children show a more precise attainment of the pitch scaling and alignment of the high targets than do Catalan and English children.

Keywords

Intonation, acquisition, English, Spanish, Catalan

1 INTRODUCTION

While perceptual data on intonational development, especially from the period from birth to up the first year is relatively abundant (see review in Vihman, Nakai & DePaolis, 2006), production data on the early childhood years, is 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, which means that they examine the overall shape and direction of the contours, rather than the phonetic detail of the different tonal targets (see review in Snow and Balog, 2002). For instance, they may concentrate in 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; Whalen, Levitt, & Wang, 1991; Snow, 2006). 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, Peppe & Goulondris, 2004). Likewise, it has also been shown that there are cross-linguistic differences in the type of contours (falling or rising) acquired first by the child (e.g. Hallé, Boysson-Bardies, & Vihman, 1991; Whalen, Levitt, & Wang, 1991).

Although such type of contour-based analysis has proven useful for describing the intonation of babbling and perhaps of one-word utterances, it falls short when it comes to describing the intonation of multi-word utterances, as it does not allow a fine-grained study of the phonetic detail of the alignment and scaling of the tonal targets with respect to the segmental structure. Thus, it becomes impossible to draw meaningful comparisons to the adult data, which is generally analysed following the conventions of the Autosegmental Metrical framework (henceforth AM). The AM framework (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Silverman et al, 1992; Ladd, 1996; Gussenhoven, 2005) proposes that the intonation of any language can be analysed as a succession of high and low tonal events, which are associated (aligned) either with the metrically prominent syllables (e.g. “ba” in “baby”) or with the right-most boundary of phrases. The first are called pitch accents or accents and can be high (H*), low (L*) or some and combinations of high and low. The latter 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 the pitch accents that precede the nuclear accent.

The AM framework has been applied to many languages and has quickly become the most widely used phonological framework for analyzing intonation as it is simple, versatile, and allows for cross-linguistic comparisons. Some recent studies that have successfully applied the AM framework to the description of early child intonation are Chen and Fikkert (2007), for Dutch; Prieto and Vanrell (2007), Prieto et al (to appear) for Catalan; and Frota and Vigário (2008), for Portuguese. Chen and Fikkert (2007) examined the productions of three children between 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 pre-nuclear accents by the time they had a vocabulary of 230 words. They argued that there exists 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 an 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 realization. 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, at 1;9 the alignment and scaling of this tonal targets are essentially correct. Prieto and Vanrell (2007) found that the 4 Catalan 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 faster 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 excessive 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; 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 this study we assume thus that duration, as determined by speech rate and by segmental composition (e.g. syllables differ in the number and type of segments; segments themselves have different intrinsic durations) will have an influence in the alignment of the tones with the segmental string. From studies on adult alignment, we know that peaks, 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) and/or speech rate. This is known as the “relative segmental anchoring hypothesis” and has been proposed for English (van Santen & Hirschberg, 1994); Dutch (Rietveld & Gussenhoven, 1995); Pisa and Bari Italian (Gili-Fivela & Savino, 2003); French (Welby & Loevenbruck, 2005); Spanish (Prieto & Torreira, 2007), among others. (But see defendants of the “segmental anchoring hypothesis”, e.g. Arvaniti, Ladd & Mennen, 1998, for Greek; Ladd et al. 1999, for English; Xu, 1998, for Mandarin; Xu & Liu, 2006, for Mandarin and English; Atterer & Ladd 2004, for German. For a summary of the tonal alignment debate see Ladd 2008: 172ff).

Data regarding the alignment of the low targets is far less conclusive, at least for English. With reference to 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. 105ff) reviews the arguments for and against this claim and concludes that this interpretation causes more problems than it solves, for then 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 shall explain in more detail in Section 1.3 below. The alignment of L in other languages such as Greek (Arvaniti, Ladd & Mennen, 1998; Ladd, Mennen & Schepman, 2000), however, has proved so robust that has been used and is still used as an argument for the stronger version of the ‘segmental anchoring hypothesis’ (Ladd 2008: 172ff).

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 Figure 1 below), which is comparable across the three languages and is abundant enough 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 (e.g. accentual range and pitch range) and alignment?; (2) If so, are developmental differences also to be found cross-linguistically?

This would be in line with the early claim that children have a reduced pitch range compared to adults (e.g. Marcos, 1987), and that this increases as the child acquires a complex tonal inventory. However, studies that examine separately falling and rising tones (e.g. Snow 1995, 1998) have shown that falling tones behave differently. The four-year-old children in Snow (1998) actually use a pitch range with falling tones which 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 and than adults. That is, both for scaling and alignment, the distance between child and adult will decrease as the child develops linguistically.

We also expect to find possible cross-linguistic differences in the age of acquisition of alignment since English, Catalan, and Spanish differ with regard to typological factors such as rhythm, syllable composition, and density of pitch accents - factors which are known to interact with the alignment of tonal targets. Previous studies of rhythm in adults (e.g. Ramus, Nespore, & Mehler, 1999; Prieto et al 2010), have found cross-linguistic rhythmic differences, leading towards establishing a rhythm typology; English would appear at one end of the scale (stress-timed, complex consonant clusters, with vowel reduction, sparse pitch accents), Spanish at the other end (syllable-timed, predominance of CV syllables, dense pitch accents), with Catalan typically placed somewhere in the middle, though perhaps leaning towards the syllable-timed end of the typological continua. Prieto et al (2010), which used the same adult participants as in the present study and a very controlled methodology, partially corroborated previous findings, as they found that Spanish and Catalan are more similar to each other, both tending towards a syllable-based rhythm than to English, which is more clearly stress-based.

The few existing cross-linguistic studies of rhythm in children (Grabe, Post, & Watson, 1999, Vihman, Nakai & DePaolis, 2006; DePaolis, R, Vihman, & Kunnari 2008; Payne et al, to appear 2012) corroborate the existence of cross-linguistic and developmental differences in the acquisition of duration. Grabe, Post, & Watson (1999) examined the variability in vocalic duration in the speech of 4-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. Vihman and collaborators (Vihman, Nakai & DePaolis, 2006) 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 et al (2010, to appear 2012), using the same participants as in the present study, concluded that language-specific characteristics co-exist with properties common to child speech irrespective of language, 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 and consonantal articulation. As these skills are mastered, cross-language similarities disappear, so that by 6 years the rhythmic characteristics are more closely aligned with those of adult speakers for each language.

Another source of variability which 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 in the acquisition of early phonological templates. For instance, in Catalan, due to the historical loss of masculine inflectional morphemes (Spanish ‘gato bonito’, Catalan ‘gat bonic’) there is a much higher number of syllables ending in a vowel (‘violí’, ‘lleó’, ‘camió’). In Spanish 95% of oxytones end in a consonant or glide (‘violín’, ‘león’, ‘camión’) (e.g. Hualde 2006) . 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 & Demuth 1998, Lleó 2006) as they do not

have to contend with complex syllabic structures and they can concentrate in increasing the number of syllables. Similar observations have been made for other languages (e.g. Vihman, DePaolis & Davis, 1998) for French and English; Vigário, Freitas & Frota, 2006, for Portuguese).

Since the 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 that 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, Stoel-Gammon & Buder, 1995; Snow, 1994; Snow 1995, 1998; Vanrell et al 2011; Prieto et al, in press) we may expect pitch scaling to be quite stable from an earlier age.

3 METHODOLOGY

We aim at comparing the intonational patterns of words controlled for number of syllables and stress position in the children's speech to the adult target provided by their mothers, both elicited with a controlled naming task. Recordings were made respectively in Cambridge, Madrid and Barcelona, on 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 children 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) who were about 2, 4 and 6 six years of age at the time of the recordings. The ages of the children were chosen so they fell into clearly differentiated developmental stages.¹

Not all the children produced enough speech and some recording had background noise or other problems. In order to have complete datasets for every language and age, we selected the best 3 recordings for every age range and language. We have thus analysed for the present study 27 children and 12 adult women who acted as controls.

Table 1

Participant's characteristics and number of tokens in each language and age group

Language	Age	Age range	Gender	Tokens
Catalan	2	2;2-2;8	F,F,M	17
	4	3;9-4;11	F,F,M	20
	6	6;4-6;10	M,M,F	39
	Adults	25-40	F	132
Spanish	2	2;3-2;7	F,F,M	20
	4	3;5-4;5	M,M,F	25
	6	5;9-6;7	F,F,M	31
	Adults	25-40	F	137
English	2	1;11-2;0	F,F,M	33
	4	3;10-4;10	F,F,F	29
	6	5;10-6;0	F,F,F	44
	Adults	25-40	F	99

The data were elicited with a naming game, based on animated pictures shown on Powerpoint slides on a laptop screen. The animations showed scenes, some with animals and some with everyday objects, that included the target word. Mothers were given written instructions explaining that they have to read a short story about a little 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 by 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 had to encourage her to try again until the child used the target word. The dialogue was modeled for her 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.

[mother] Can you find it? There! Well done

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, which were selected so they were imageable and familiar to young children. We used words such as ‘train’, ‘tren’ and ‘tren’ (monosyllabic), ‘balloon’, ‘camión’, ‘camió’ (WS); ‘baby’, ‘mono’, ‘mono’ (SW), etc. (See word list in Appendix). We also included in the analyses 5 pairs of monosyllabic and disyllabic words (e.g. ‘key’ and ‘monkey’), which were recorded for a different experiment. The S and WS categories are larger than the rest but this is in line with their higher frequency of occurrence in child speech in all three languages under study.

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 completed the gaps 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 should be easy to pronounce and as comparable as possible across languages; open and closed syllables were balanced across languages.

The age of the children posed some methodological problems, since we had to rely exclusively on words that were both imageable and maximally familiar to young children and also, as far as possible, easy to pronounce for the younger children. Because of the tendency to simplify complex codas in young children’s speech (Gerken 1994; Vihman, DePaolis, & Davis 1998), we mostly used open syllables. And because of the possible effects of stressed syllable type in alignment (e.g. Schepman, Lickley & Ladd 2006; Prieto & Torreira 2007), we maintained a balance of open and closed syllables in the three languages across the whole database. We have between 6 and 8 target words with closed stressed syllables per language:

English: 7 words (helicopter, violin, balloon, guitar, monkey, moon, train

Spanish : 8 words (helicoptero, elefante, violin, pantalon, caracol, leon, camion, tren)

Catalan: 6 words (helicopter, elefant, cocodril, pantalons, taxi [tak-si], tren

The goal was to obtain target words that were comparable across the three languages and, to the extent to which this is possible, to the adult target.

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 which has different versions for different languages. For Spanish, we used Sp_ToBI, the model proposed in Estebas & Prieto (2008, 2010). For Catalan, we used Cat_ToBI (Prieto et al., 2009). For English, we based our transcription on Gussenhoven's proposal for Dutch (Gussenhoven 2005) because the intonational phonology of Dutch is very similar to that of English (see Gussenhoven 2004: 297ff, for a description of the intonational phonology of English). One significant departure from the original ToBI model proposed by Gussenhoven (2004, 2005) and 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 intonational phrase (IP) boundary tones.

We have chosen for study an intonation contour which is comparable across the three languages and which occurs frequently in our database. This contour is characterized by an optional rise (onglide) from the beginning of the utterance to the high target, followed by a fall, as shown below:

Figure 1. Schematic representation of two phonetic variants of the target contour

In Spanish and Catalan the initial rising onglide (represented by the L+H* accent) is generally present whereas in English it may not appear. 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” that do not provide enough segmental material to realize the initial rising movement. However, even in contexts where there is enough voiced material, the rising onglide can still be omitted. The difficulty in distinguishing between L+H* and H* has been noted repeatedly in the literature on English intonation. Pitrelli et al. (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 both categories in their analysis of the data. Ladd (1996, p. 84ff) comments that L+H* “[o]n a phrase initial accented syllable, L+H* and H* can be difficult to distinguish”. Gussenhoven (2004: 298) analyses the L as an optional low onset that aligns right at the start of the accented syllable.

In our data we found another pattern, one in which the H* tone is sustained and stretched on to a plateau, often with a very level pitch range, as represented in the Figure:

Figure 2. Schematic representation of L+H* with a plateau

This configuration is extremely frequent in the child data, especially in 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

undershot the rising and falling movements. Thus, we may consider it as an underlying (L)+H* pitch accent, although in order to validate this hypothesis, we will have to undertake further testing.

4 ANALYSIS

The recordings were digitized and processed using Praat by English, Catalan and Spanish 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 labeled 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 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 et al, 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 which can trigger downstep (lowering in scaling) of 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

H1	f0 peak location in Type 1 accent
H2	f0 peak location in the Type 2 accent
H*	f0 peak location in the Type 3 accent
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, Mennen, & Schepman, 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. Microprosodic effects (such as the dip produced by nasal segments) were disregarded.

An example of the measurement points which illustrate the labeling 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 4-years-old and one 2-years-old (realized as ‘monnies’ [ˈmʌnɪs]).

Figure 3. On the top row, from left to right pitch traces for ['mɐni] and ['mɐnis] with Type 1 and Type 2 accents. On the second row, pitch trace of ['mɐni] with Type 3 accent.

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 phrase-final position and were obtained with a task that elicited narrow focus on the nuclear syllable. From observations and from descriptions in the literature, in Spanish and in Catalan we expect 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 postonic

syllable; Prieto et al, 1995, for Spanish; Estebas-Vilaplana, 2000 for Catalan). We may expect that in syllables with sonorant codas, the H* may be aligned earlier in the syllable (Prieto and Torreira 2007). In English, the H* peak is aligned with the stressed syllable, and the L, when present, with the onset of the stressed syllable (Gussenhoven 2004: 297ff).

Three different measures of L and H alignment were used for statistical exploration: (i) L to Syllable Onset, (ii) H* to Syllable Onset, (iii) H* to Syllable End. We expect that (i) will be appropriate for the data because the leading L tone is anchored at the onset of the stressed syllable. Both (ii) and (iii) have been used in the literature on alignment. For instance, Prieto et al (1995) used both H* to Syllable Onset and H* to Syllable End. Xu (1998) and Xu and Liu (2006) proposed that syllabic onsets provide a more robust anchoring point for tonal movements. Schepman, Lickley, & Ladd (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 and 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 have decided to use both measures of H* alignment, H* to Syllable Onset and H* to Syllable End.

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

(4) Stressed syllable type labels

(P)V(P) no sonorant onset or coda – rime only

SV sonorant onset

(C)VS sonorant coda

SVS both sonorant onset and coda

5 RESULTS

The database contained 3120 measurement points; five measurement points for each of the 624 tokens. 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. Prior to the analyses, we explored these data using boxplots. Outliers and extreme values were identified and re-measured by hand, but 7 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 the following Figure.

Figure 4. Pie chart showing the distribution of Type 3 accents according to age group as a percentage of the total

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, only produced 8% of plateaux. These results are in line with previous findings about 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 totally level.

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

In order to compare the children productions to ‘normal’ adult speech, a new dataset of Adult Directed Speech (ADS) was collected at a later stage. For each language, 3 adult women were recorded performing the same tasks with each other or with the researchers. The new ADS dataset has 242 cases, thus making a total of 467 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% (many studies of child pitch range call this measure “accent range”, but it is in fact the H* to L% distance what they measure; see Snow 1995). Figure 5 shows the pitch range for each language and age group.

Figure 5. For each age group (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, two-year-olds have a pitch range of 7 st, four-year-olds of 6.2 st, six-year-olds of 4.3 st. Adults using Adult Directed Speech (ADS) have a pitch range of 5 st. An ANOVA with dependent variable Pitch Range and with between-subjects factors Age and Language confirm statistically significant effects of Age ($F(3, 455)=16.581, p=0.001, \text{partial eta-squared}=0.099$), Language $F(2,$

455)=9.594, $p=0.000$, partial eta-squared=0.040), and a significant Age by Language interaction $F(6, 455)=4.909$, $p=0.000$, partial eta-squared=0.061).

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, to Language, and to Stress (Oxytones, and Non-oxytones). 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 on the literature (e.g. Allen & Hawkins 1978, 1980; Snow 1995, 1998), that young children have to learn how to control pitch properly while adults and more linguistically mature children show a robust scaling of the pitch accent which is quite constant across different stress patterns. Adults can achieve the appropriate scaling even in word and phrase final position, when there is strong pressure of the upcoming tonal boundary. Allen & Hawkins (1978) analysed monosyllabic and disyllabic words from a corpus of natural speech produce by three children at age 3 and concluded that children produce 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 to 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. "bottle" and "shoe").

Figure 6 shows the pitch range for each age group and for each stress pattern.

Figure 6. Mean f0 difference in st 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.

An ANOVA with dependent variable Pitch range and with between-subjects factors Word Pattern, Language and Age, confirm statistically significant effects of Age ($F(3, 443)=17.646$, $p=0.000$, partial eta-squared=0.107), Language $F(2, 443)=11.463$, $p=0.000$, partial eta-squared=0.049), and a significant Age by Language interaction $F(6, 443)=5.278$, $p=0.000$, partial eta-squared=0.009). Stress, on the other hand, is not significant and there are no significant interactions with Age or Language.

5.2 Alignment

We analysed the alignment of the H* peak in the 387 tokens produced with Type 1 and Type 2 accents (see Figures 1 and 3). 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 below 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.

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.

As we see in the error bars in Figure 7, even the youngest children show a remarkable precise alignment of the L target.

In order to further explore the data, we ran a series of ANCOVAS for each alignment measure with between-subjects factors Age and Language and with covariate factor Syllable Duration. The analyses confirm statistically significant effects of Syllable Duration ($F(1,261)=16.504$, $p=0.000$, 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), but extremely small. Catalan 2 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. These differences are similar to those between the Spanish age groups, which are not statistically significant.

As for the alignment of the H* target, we see in the error bars in Figure 8 below that adults are much closer to the target than children are. At a glance, we appreciate as well cross-linguistic differences in both measures of alignment for H*.

Figure 8. Distance (in seconds) from the H* 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 English, Spanish, and Catalan.

Figure 9. Distance (in seconds) from the H* target to the syllable end (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 English, Spanish, and Catalan.

Spanish two-year-olds are much closer to the adult target than English and Catalan. As Catalan and English children grow older, they align their H* increasingly closer to the end of the syllable and, eventually, Catalan six-year-olds catch up with the Spanish children. As for the adult data, Catalan and Spanish adults align their H* target slightly closer to the end of the syllable than English speakers do. We notice as well much more variability, especially for the younger children, in the alignment of H* to Syll End (in Figure 8) than on the alignment of H* to Syll Onset (in Figure 9). We hypothesize 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 then 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 (H^* to Syll Onset $F(1, 355)=68.486$, $p=0.000$, partial eta-squared=0.162), Age H^* to Syll Onset $F(3, 355)=10.642$, $p=0.000$, 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 significance level.

As for the alignment of H* to Syllable End, there are strong significant effects of Syllable Duration ($F(1, 355) = 864.775, p=0.000$, partial eta-squared=0.709), and weaker effects of Age ($F(3, 355) = 8.488, p=0.000$, 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 H* peak alignment, we see that Duration is the only variable that has strong effects (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.

We have so far found very small (although statistically significant) effects of the independent variable Stress on the alignment of H* to Syll End. To examine this further, in the following round of analyses, we break up the data according to the stressed pattern of the target words. Figure 9 shows the alignment of H* to Syll Onset and to Syll End for each age group and for each stress pattern in relation to syllabic duration. Figure 10 shows the alignment of H* to Syll Onset and to Syll End for each age group and for each stress pattern in relation to syllabic duration.

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. The two panels on the first row show correspond to ages 2 and 4; the two panels on the second row to ages 6 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. The two panels on the first row show correspond to ages 2 and 4; the two panels on the second row to ages 6 and adults. The target words are separated in oxytones (a circle), paroxytones (a triangle) and proparoxytones (a star).

As we see in the panels in Figure 10, and in Figure 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 it is H* to Syllable Onset (Pearson=-0.556, $p<0.001$).

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 in the alignment of oxytones, paroxytones and proparoxytones than younger children do. For the two and four-year-olds, the difference between

oxytones on the one hand and paroxytones and proparoxytones on the other is of the order of 100 ms; for adults and six-year-olds it is about 50 ms. Younger show similar patterns of alignment to adults and older children, but with more variability.

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 with covariate factor Syllable Duration. For the alignment of H to Syll Onset, the analyses confirm statistically significant effects of Syllable Duration ($F(1, 313)=122.288, p=0.000, \text{partial } \eta^2=0.281$), Stress ($F(2, 313)=35.360, p=0.000, \text{partial } \eta^2=0.184$), and Age ($F(3, 313)=4.457, p=0.004, \text{partial } \eta^2=0.041$). The interaction of Age by Stress is also significant ($F(6, 313)=2.389, p=0.028, \text{partial } \eta^2=0.044$).

As for the alignment of H to Syll End, we find significant effects of Syllable Duration ($F(1, 313)=449.004, p=0.000, \text{partial } \eta^2=0.589$), Stress ($F(2, 313)=45.441, p=0.000, \text{partial } \eta^2=0.225$), Age ($F(3, 313)=2.460, p=0.000, \text{partial } \eta^2=0.023$), and of the Age by Stress interaction ($F(1, 313)=2.851, p=0.010, \text{partial } \eta^2=0.052$).

The stress pattern of the word (Stress) has a larger effect on the alignment of H^* to Syll End, ($\text{partial } \eta^2=0.225$), but not on the alignment of H^* to Syll Onset ($\text{partial } \eta^2=0.184$), and this effect is larger in older children and in adults. 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 correlates with syllabic duration. Effectively, the accented syllable of oxytone words is the last one in the word and, in this data, also the last one 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 syllabic duration may in turn influence the alignment of the H^* target.

In order to rule out potential structural effects, 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), 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 ($F(1, 299) = 348.465$, $p=0.000$, partial eta-squared=0.567), H* to Syll Onset ($F(1, 299) = 155.225$, $p=0.000$, 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 extremely small (L to Syll Onset ($F(1, 299) = 7.752$, $p=0.006$, partial eta-squared=0.028).

The stress pattern of the word (Stress: oxytone, paroxytone, proparoxytone) has significant though very modest effects on the alignment of the H* target (H* to Syll Onset $F(2, 299) = 23.079$, $p=0.000$, partial eta-squared=0.139; H* to Syll End ($F(2, 299) = 28.677$, $p=0.000$, partial eta-squared=0.167), but not on the alignment of the L target. The type of Syllable Structure does not have any significant effect on the three alignment measures, not 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.

As shown by the effect sizes, the duration of the stressed syllable has a smaller effect 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, but this is very small (partial eta-squared=0.184 and partial eta-squared=0.225). For both measures, Age is statistically significant, although it has very small effects on both measures of H* alignment (partial eta-squared=0.041 and partial eta-squared=0.052). Finally, for both measures, there is a statistically significant interaction of Stress and Age, also with very small effects (partial eta-squared=0.044 partial eta-squared=0.052).

6 CONCLUSION

Performing a fine-grained quantitative analysis of the scaling and alignment of tonal targets based on the Autosegmental Metrical method, we have found that the younger children can control the mechanism of 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 increasingly less, and these 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. Since the English adult language has relatively wide pitch range excursions, they are not far from the mark. Catalan and Spanish children, on the other hand, have a longer way to go, as the adult language uses a relatively narrow pitch range –for these children, achieving adult-like pitch scaling implies narrowing their pitch range. We have also noticed that six-year-olds have a pitch range similar to that of adults; six-years old are thus closer to adult-directed language.

In line with previous findings in English (Snow 1995, 1998) we found that children control pitch so that they produce a similar pitch range for paroxytones ('bottle') than for oxytones ('shoe'), just as adults do. In our data, we found significant effects of Age and an Age by Language interaction, but not effects of Stress pattern. The fact of whether the word is an oxytone or not, does not have significant effects on pitch range. Children show a very good control of pitch although they may have difficulty controlling the timing aspects of prosody, as has been reported in the literature (from Allen & Hawkins 1978; to Payne, to appear 2012). Indeed, the children in the present study were also analysed doing a different task by Payne et al (2011, to appear 2012) who report excessive lengthening of final syllables.

Despite their lack of mastery in the control of syllabic durations, even our youngest children produced adult-like alignment of the L target right from the start. And although their alignment of the H* target is not as accurate as that of adults, this is because of the effects of syllabic 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) in just one syllable and as consequence of this tonal

crowding, the H* target is realised earlier into the syllable. But the fact that children have acquired at such a young age the basic alignment constraint, that is, the constraint to realise both tonal targets in one syllable, is in itself remarkable.

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 have found significant Age by Language interactions, although the differences between the child speech and the adult speech were very small, as attested by the reduced effect sizes that we found. Young Spanish children show a more precise command of the alignment of the high targets than Catalan and English children, but it is true that the Spanish children were slightly younger and this may have had an effect, especially at Age 2, which is time of rapid phonological development. We thus, interpret these apparent cross-linguistic differences with caution. Both in alignment and in pitch range, the important fact is that younger children are getting increasingly closer to the adult target in the language they are learning.

FOOTNOTES

¹ As noted by the reviewers, the English children were slightly younger overall than the Catalan and Spanish children. This is more may cause more difficulties for comparing data at Age 2, an age of rapid development, than for Ages 4 and 6, and this means that we will have to use some caution in interpreting any possible cross-linguistic differences at this age.

REFERENCES

- Allen, G.D. & Hawkins, S. (1978). The development of phonological rhythm. In A. Bell and J.B. Hooper (eds) *Syllables and Segments*. Amsterdam: North-Holland Publishing.
- Allen, G.D. & Hawkins S. (1980). Phonological rhythm: definition and development. *Child Phonology*. Volume I: *Production*. Academic Press.

- Arvaniti, A., Ladd, D.R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26 .1:3-25.
- Astruc, L., Payne, E., Post, B., Prieto, P., Vanrell, M.M. (2010). Word prosody in early child Catalan, Spanish and English. *Proceedings of Speech Prosody*, May 2010.
- Atterer, M. & Ladd, D.R. (2004). On the phonetics and phonology of 'segmental anchoring of F0: evidence from German'. *Journal of Phonetics* 32.2:177-197.
- Beckman, M. & Pierrehumbert, J. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3: 15-70.
- Boersma, P. & Weenink, D. (2004). *Praat: doing phonetics by computer (Version 4.2.31)* {Computer program}. <<http://www.praat.org/>> (Retrieved April 30).
- Chen, A. & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of ICPhS 2007*, 315-320.
- Crystal, D. (1979). Prosodic development. In P.J. Fletcher and M.A. Garman (Eds), *Language acquisition*. (pp.33-48). Cambridge: CUP. 2nd edn, 1986, (174-197).
- DePaolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study'. *Journal of Phonetics* 36:406-422.
- Estebas Vilaplana, E. (2009). The use and realisation of accentual focus in Catalan with a comparison to English. München: Lincom Europa.
- Estebas, E. & Prieto, P. (2008). La notación prosódica del español: una revisión del Sp_ToBI. *Estudios de Fonética Experimental* XVII, 263-283.
- Estebas-Vilaplana, E. & Prieto, P. (2010). Castilian Spanish intonation. In: Prieto, P.; Roseano, P. (coords.) (2010). *Transcription of Intonation of the Spanish Language*. München: Lincom Europa, 17-48
- Frota, S. & Vigário, M. (2008) .The intonation of one-word and first two-word utterances in European Portuguese. Paper presented at the *XI International Conference for the Study of Child Language* (IASCL), Edinburgh, August 1-4, 2008.
- Galligan, R. (1986). Intonation with single words: purposive and grammatical use. *Journal of Child Language* 14, 1-21.

- Gerken, L.A. (1994). A metrical template account of children's weak syllable omissions. *Journal of Child Language*, 21(3), 565-584.
- Gili-Fivela, B. & Savino, M. (2003). Segments, syllables and tonal alignment: a study on two varieties of Italian. In: M.J. Sole', D. Recasens and J. Romero (Eds.) *Proceedings of the XVth International Congress of Phonetic Sciences* (pp.2933-2936). Causal Productions: Barcelona.
- Grabe, E. (1998). *Intonational phonology: English and German*. Doctoral dissertation. Max-Planck-Institut for Psycholinguistics and University of Nijmegen.
- Grabe, E., Post, B. & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. *Proceedings ICPHS 1999*, San Francisco, pp.1201-1204.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: CUP.
- Gussenhoven, C. (2005). Transcription of Dutch Intonation. In Sun-Ah Jun (Ed.) *Prosodic typology: The phonology of intonation and phrasing* (pp.118-145). Oxford University Press,.
- Hallé, P., de Boysson-Bardies, B., & Vihman, M. (1991). Beginnings of prosodic organization: rules and intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech* 34.4:299-318.
- Hualde, J.I. (2005). *The sounds of Spanish*. Cambridge: CUP.
- Kehoe, M., C. K. Stoel-Gammon & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research* 38:338-350.
- Kent, R.D. & Murray, A.D. (1982) Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the acoustic society of America* 72:353-363.
- Knight, R.A and. Nolan, F. (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association* 36.1:21-38.
- Ladd, D.R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladd, D.R. (2008) *Intonational phonology*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.
- Ladd, D.R., Faulkner, D., Faulkner, H. & Schepman, A. (1999). Constant "segmental anchoring" of f0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106:1543-1554.

- Ladd, D.R., Mennen, I. & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107.5:2685-2696.
- Lleó, C. (2006). The acquisition of prosodic word structures in Spanish by monolingual and Spanish-German bilingual children. *Language and Speech*, 49, 207-231.
- Lleó, C. & Demuth, K. (1999). Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages. *Proceedings 23rd BUCLD*, 407-418.
- Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language* 14.2:255-68.
- Payne, E.; Post, B.; Astruc, L.; Prieto, P.; & Vanrell, M.M. (to appear 2012). Measuring rhythm in child-mother interaction. *Language and Speech*.
- Payne, E., Post, B., Astruc L., Prieto, P., & Vanrell, M.M. (2010) Rhythmic modification in Child Directed Speech. In Michela Russo (Ed) Atti del convegno di Prosodia. Gli universali prosodici: confronto e ricerche sulla modellizzazione rítmica e sulle tipologie ritmiche. Roma: Aracne Biblioteca di Linguistica.
- Pierrehumbert, J. (1984). *The phonetics and phonology of English intonation*. Doctoral Dissertation. MIT.
- Pitrelli, J.F., Beckman, M.E., and Hirschberg, J. (1994). Evaluation of prosodic transcription labelling reliability in the ToBI framework. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP-1994)*, pp.123-126.
- Prieto, P., van Santen, J., Hirschberg, J. (1995) Tonal alignment patterns in Spanish. *Journal of Phonetics*. 23: 429-51.
- Prieto, P. (2006). The relevance of metrical information in early prosodic word acquisition: A comparison of Catalan and Spanish. *Language and Speech*, 49, 231-259.
- Prieto, P. & Vanrell, M.M. (2007). Early intonational development in Catalan. *Proceedings of ICPHS XVI* (pp. 309-314). Trouvain, J. and W. J. Barry (Eds.) Pirrot GmbH: Dudweiler.
- Prieto, P. & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35.4:473-500.

- Prieto, P., Aguilar Cuevas, L., Mascaró i Pons, I., Torres-Tamarit, F., & Vanrell, M.M. (2009). 'L'etiquetage prosòdic Cat_ToBI'. *Estudios de Fonética Experimental* 18: 287-309.
- Prieto, P., Estrella, A., Thorson, J., Vanrell, M.M. (to appear). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*.
- Prieto, P., Vanrell, M.M., Astruc, L., Payne, E., & Post, E. (2010) Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English, and Spanish. *Proceedings of Speech Prosody*.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73.3:265-292.
- Rietveld, T. & Gussenhoven, C. (1995). Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* 23.4:375-385.
- Santen, J. P. H. & Hirschberg, J. (1994). Segmental effects on timing and height of pitch contours. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP-1994)*, pp. 719-722.
- Schepman, A., Lickley, R. & Ladd, D.R. (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34:1-28.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). 'TOBI: A standard for labeling English prosody.' *Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP-1992)*, pp.867-870.
- Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research* 37:831-840.
- Snow, D. (1995). Formal regularity of the falling tone in children's early meaningful speech. *Journal of Phonetics* 23:387-405.
- Snow, D. (1998). Children's imitations of intonation contours: are rising tones more difficult than falling tone? *Journal of Speech, Language, and Hearing Research* 41, 576-587.
- Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development* 77.2:281-296.

- Snow, D. & Balog, H.L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua* 112, 1025-1058.
- Vanrell, M. M.; Prieto, P.; Astruc, L.; Payne, E.; Post, B. (2011). Adquisició dels patrons d'alineació i de camp tonal en català i espanyol. *Quaderns de Filologia*
- Vigário, M., Freitas, M.J., & Frota, S. (2006). Grammar and frequency effects in the acquisition of prosodic words in European Portuguese. *Language and Speech*, 49, 175-2003.
- Vihman, M. M., Nakai, S., & DePaolis, R. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen and C. Best (Eds.) *Laboratory Phonology* 8 (pp. 341-366). Mouton de Gruyter: New York.
- Vihman, M.M., DePaolis, R., & 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.
- Welby, P. & Loevenbruck, H. (2005). Segmental "anchorage" and the French late rise. In: *Proceedings of Interspeech* (pp. 2369–2372).
- Wells, B., Peppé, S., & Goulandris, A. (2004). Intonation development from five to thirteen. *Journal of Child Language* 31:749-778.
- Whalen, D.H., Levitt, A.G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language* 18, 5001-516.
- Wilson, M.D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20 (1), 6-11.
<http://www.psych.rl.ac.uk>
- Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.
- Xu, Y. & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: toward an integrated model. *Italian Journal of Linguistics* 18.1: 125–159.

APPENDIX

Stress pattern	English	Spanish	Catalan
S	Sun ['sʌn] Train ['treɪn] Bee ['bi:] Moon ['mu:n] Tea ['ti:] Knee ['ni:] Key ['ki:s]	Sol ['sol] Tren ['tren] Pie ['pie] Bus ['bus] Sí ['si] Be ['be] Pe ['pe]	Sol ['sol] Tren ['tren] Peu ['peu] Bus ['bus] Sí ['si] Be ['be] Pe ['pe]
SW	Angel ['eɪndʒ.əl] Baby ['beɪ.bi] Lion ['laɪ.ən] Monkey ['mʌŋ.ki] Potty ['pɒ.ti] * Daddy ['dæ.di] * Mummy ['mʌ.mi] * Money ['mʌ.ni] *	Nena ['ne.na] Luna ['lu.na] Mono ['mo.no] Llave ['ʎa.βe] Papi ['pa.pi] Mami ['ma.mi] Globo ['glo.βo] Bibe ['bi.βe]	Nena ['nɛ.nə] Lluna ['ʎu.nə] Mono ['mo.nu] Taxi ['tak.si] Papa ['pa.pə] Mama ['ma.mə] Globus ['gɫɔ.βus] Bibe ['bi.βə]
WS	Guitar [ɡɪ.'tɑ:] Balloon [bə.'lu:n] Giraffe [dʒə.'raf]	Bebé [be.'βe] Camión [ka.'mjɔn] León [le.'on]	Bebè [bə.'βɛ] Camió [ka.'mjo] Lleó [ʎə.'o]
WSW	Potato [pə.'teɪ.tou] Banana [bə.'nɑ:n.ə] Tomato [tə.'mɑ:tou] Pajamas [pə.'dʒɑ:məs]	Zapato [θa.'pa.to] Pelota [pe.lo.ta] Gallina [ga.'ʎi.na] Pijama [pi.xa.ma]	Sabata [sa.'βə.tə] Pilota [pi.'ɫə.tə] Gallina [ga.'ʎi.nə] Pijama [pi.'ʒa.mə]
SWW	Elephant ['el.ɪf.ənt] * Crocodile ['krɒk.ə.daɪl] * Melanie ['mæl.an.i] * Butterfly ['bʌt.ə.flai] *	Águila ['a.ɣi.la] Música ['mu.si.ka] Mélanie ['me.la.ni]	Àliga ['a.li.ɣə] Música ['mu.si.kə] Mèlanie ['me.la.ni]
WWS	Cockatoo [kək.ə.'tu:] Kangaroo [kæŋ.gə.'ru:] Violin [vaɪə.'lɪn]	Caracol [ka.ra.'kol] Pantalón [pan.ta.'lon] Violín [bi.o.'lin]	Cocodril [ku.ku.'ðril] Pantaló(ns) [pən.tə.'lo(ns)] Elefant [ə.lə.'fan] Ratolí [ra.tu.'li]
SWSW	Caterpillar ['kæt.ə.pɪl.ə] * Watermelon ['wɔ:t.ə.,mel.ən] * Helicopter ['hel.ɪ.kəpt.ə] *	Mariposa [ma.ri.'po.sa] Elefante [e.le.'fan.te] Cocodrilo [ko.ko.'ðri.lo]	Papallona [pə.pə.'llo.nə] Helicòpter [ə.li.'kɔp.tər] Hipopòtam [i.pu.'pɔ.təm]
SWSWW	Hippopotamus [ˌhɪp.ə.'pɒt.əm.əs] *	Hipopótamo [i.po.'po.ta.mo] Helicóptero [e.li.'kop.te.ro]	

* We assume ambisyllabicity in the starred cases as this facilitates the comparison with Spanish and Catalan. An alternative syllabification would be ['pɒt i], with the /t/ grouped with the stressed syllable.