Acquisition of tonal targets in Catalan, Spanish, and English[*]

Lluïsa Astruc[1,2], Pilar Prieto[3,4], Elinor Payne[5], Brechtje Post[1], Maria del Mar Vanrell[6]

Cambridge University[1], The Open University[2], Oxford University[5], ICREA[3], Universitat Pompeu Fabra[4,] Universitat Autònoma de Barcelona[6]

This study analyzes the distribution, scaling and alignment of low and high targets in the productions 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 and they still lengthen excessively final syllables. 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 influences the acquisition of tonal targets; young Spanish children show a more precise alignment of pitch scaling and of the alignment of the high targets than Catalan and English children.

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 et al 2006), production data is very scarce, especially data on the early childhood years. 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 (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 they acquire falling tones first and final rising tones much later (Snow 1998; Wells, Peppe and Goulandris 2004).

Although the contour-based approach has proven useful for describing the intonation of babbling and perhaps of one-word utterances but it falls short when it comes to describing the intonation of multi-word utterances. This approach, for instance, 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 described in the literature, which is generally analysed following the conventions of the Autosegmental Metrical framework[1] (henceforth AM). The AM (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Ladd 1996, Gussenhoven 2005) 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, in press), 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[2] 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 exists a correlation between the expansion of the vocabulary and the expansion of the tonal inventory. Frota and

[1]   The AM (Pierrehumbert 1980, Ladd 1996) 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%).

[2]   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.

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 6 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.

We assume that prosodic factors such as duration, as determined by speech rate and by segmental composition (e.g. syllables differ in the number and type or 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, IP) and/or speech rate. This is known as the "relative segmental anchoring hypothesis" and has been convincingly proposed for English (van Santen and Hirschberg, 1994); Dutch (Rietveld and Gussenhoven 1995); Pisa and Bari Italian (Gili-Fivela and Savino 2003); French (Welby and Loevenbruck 2005); Spanish (Prieto and Torreira 2007), among others[3].

In sum, work on tonal alignment in different languages has shown that the position of the peak changes across syllable structures, although it remains consistently aligned in relation to the stressed syllable, in the way stipulated in the phonological rules of each language. Data regarding the alignment of the low targets is far less conclusive. Ladd (1996: 105ff) reviews the arguments for and against treating valleys (low points between peaks) as "sagging transitions", that is mere interpolations between two tonal targets (Pierrehumbert's original position, e.g. 1984). He claims 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.

Previous studies agree in that children produce excessive syllabic duration and that they need time to learn to produce appropriate contrasts in syllable length (stressed-unstressed, final-non final). By 2, children control some phonetic correlates of stress (pitch, intensity, duration) in familiar words (Klein 1984; Kehoe, Stoel-Gammon and Buder 1995). Kehoe and collaborators tested English children 1;8 to 3;0 and found that the older children produced appropriate stress-unstressed durational contrasts (Kehoe, Stoel-Gammon and Budder 1995, Kehoe and Stoel-Gammon 1997) . The age of acquisition of stress may be be even earlier in other languages (e.g. Hochberg 1987 in Spanish; DePaolis, Vihman and Kunnari 2008). Snow (1994) found that children start to control final lengthening after the onset of the multi-word stage (1;5-2;0) but they may experience a regression a few months later (see also Snow 2006). Snow concludes that "the timing aspects of prosody develop more slowly than intonation" (Snow 1994).

2   GENERAL GOALS AND RESEARCH QUESTIONS

In this study we examine the scaling and tonal alignment of the valley and peak and 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?

Previous studies claim that young children to show a narrower accent range than adults, overall. This would be in line with (e.g. Crystal 1978, Marcos 1987) that claim that children have a reduced pitch range compared to adults, and that this increases as the child acquire a complex tonal inventory. However, studies that differenciate between 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 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 crosslinguistic 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 et al 1999, Prieto et al (2008), 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 is typically placed somewhere in the middle, though perhaps leaning towards the syllable-timed end of the typological continua. Prieto et al (2008), 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

---

[3]   But see defendants of the "segmental anchoring hypothesis", e.g. (Arvaniti, Ladd and Mennen 1998, for Greek; Ladd et al. 1999, for English; Xu 1998 for Mandarin, Xu and Liu 2006, for Mandarin and English; Atterer and Ladd 2004, for German.

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 (Allen and Hawkins 1978, Grabe and Post 1999, Payne et al, submitted) corroborate the existence of crosslinguistic and developmental differences. Therefore 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 and scaling earlier as well.

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.

1.1 Participants

We recorded 36 children and their mothers. The children (12 English, 12 Catalan, and 12 Spanish) were about two, four 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. Of those, we have analysed for the present study 27 children (3 for every age range and language) and 12 adults who acted as controls.

1.2 Experimental material

We controlled the prosodic composition of the target words, using words with different stress patterns, from S to SWSW, 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). Words should be easy to pronounce and as comparable as possible across languages. Finally, V, CV and CVC, CCVC syllables were balanced across languages. The goal was to elicit utterances that were comparable across the three languages and, to the extent to which this is possible, to the adult target. Children delete or modify segments and even whole syllables. For instance, young children may realise "banana" (WSW) as "nana" (SW), omitting the initial unstressed syllable. The omission of unstressed syllables is called "truncation" and has been described frequently in the literature (e.g. Gerken 1994, 1996; Fikkert 1994; Archibald 1995; Demuth 1996). Truncation is especially frequent in stress-timed languages; children learning stress-timed languages such as English, German or Dutch produce truncated words until much later that children learning syllable-timed languages such as Spanish and Catalan (Lleó and Demuth 1999, Lleó 2002).

1.3 Elicitation method

The data were elicited with a naming game, based on short, animated clips, shown on Powerpoint slides on a laptop screen. Mothers were given written instructions. They have to read a short story about a little fairy called Melanie who was looking for some objects and animals. The animations showed scenes, some with animals and some with everyday objects, that included the target word. 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 "ballon", 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

1.3 Intonational contour

For the intonational analysis of the data, we followed the labeling 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 and Prieto (2009). For Catalan, we used Cat_ToBI (Prieto et al. 2009). For English, we based our transcription on Gussenhoven 2005) . One significant

departure from the original ToBI model that we have adopted in our analysis is the lack of phrase accents (L- or H-) associated to the end of phrases; another the use of the label M% to mark middle-level tonal boundaries. Thus, our phrases end with just intonational phrase (IP) boundary tones, which can be L%, M%, or H%, or potentially, a combination of those (as proposed in Cat_ToBI, Prieto et al 2009).

We have chosen for study an intonation contour which is comparable across the three languages and which is also relatively frequent 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*

{   Me   la   nie }   {   train   }
L+H*   L%   (L)+H*   L%
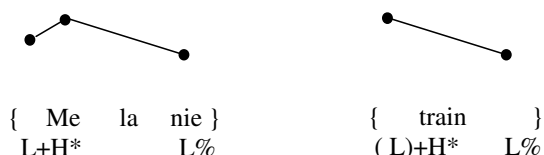
In Spanish and Catalan the initial rising onglide (represented by the L+H* accent) is generally present whereas in English it may not appear. It seems to be optional, in free alternation, as maintained by Grabe (1998: Chapter 4). For instance, the rising onglide may be absent in short words with voiceless material on the stressesed syllable like "<u>tea</u>", or "cocka<u>too</u>" 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. 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: 84ff) comments that L+H* "[o]n a phrase initial accented syllable, L+H* and H* can be difficult to distinguish". Gussenhoven (2004: 298ff) comments that he finds it hard to discern any meaning difference between L+H* and H. We also follow this interpretation, namely, that they are not two separate phonological categories, although we leave open for further examination the question of which one of them is a minor variant, an allotone, of the other.

In our data we found yet another variant of the above; one in which the H* tone is sustained and stretched on to a plateau, as represented in the Figure:

Figure 2   *Schematic representation of a third phonetic variant of the target contour*

{   Me   la   nie }
L+H*   L%
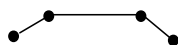
This configuration is extremely f   pecially in the younger children, although the mothers also use it occasionally. We also consider it as a variant of the L+H* pitch accent. In section 5 below we describe its use and distribution

3   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 following the loudness level (1 quiet or shy, 2 normal, 3 loud or emphatic),. We also coded the prosodic pattern of each word as was actually realised by the child, since truncation is relatively frequent in early child speech. We also labelled the indefinite article and we coded it in the database. The indefinite article ("un", "una") is stressed in Catalan and Spanish (Hualde 2005: 234) and 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 first, second and third author, 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 and 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. 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 word "Melanie".

Figure 3   *On the top row, to from left to right pitch traces corresponding to the word "Melanie" with Accent Type 1 and Type 2. On the second row, pitch trace corresponding to "Melanie" with Type 3*
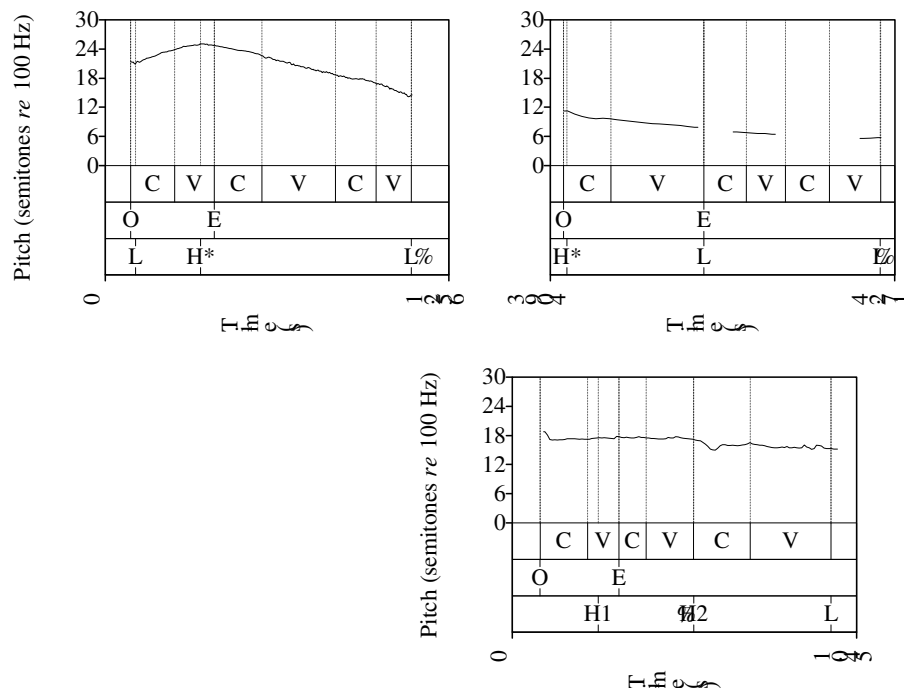

Three different measures of H and L alignment were used for statistical exploration: (i) L to Syllable Onset, (ii) H* to Syllable Onset, (iii) H* to Syllable End. We based our analysis on these measures following recent methodological arguments and our previous observations of the data. We expected that the two latter measures would provide a better fit to the data, since it has been shown that "the more distant the landmark, the greater the variance, and the greater the likelihood of uninformative correlations" (see Schepman et al 2006, Atterer and Ladd 2004), Prieto and Torreira 2007).

Pitch scaling (Pitch range) was measured in this way:used: HminL% or the distance between H* and the lowest value at the end of the utterance.

5 RESULTS

The data base contained 3120 measurement points; five measurement points for each of the 624 tokens. Prior to the analyses, we explored these data using boxplots. Outliers and extreme values were identified and re-measured by hand, but some of them had to be discarded. We did not include in the analyses tokens with two accents, where the second one was downstepped to the first, as this would be likely to affect the alignment of both. The final dataset has 584 token in total. Of those, 395 tokens were produced with a falling tone, with or without an initial L onglide and 229 with an extended level tone. The level tones were mostly produced by the younger children, as we see from the following Figure.

Figure 4   *Pie chart showing the distribution of level tones according to age group as a percentage of the total*
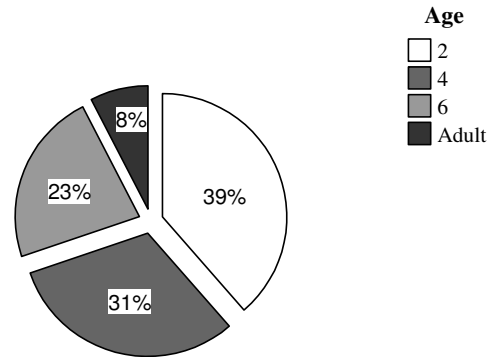

Two-year olds produced 39% of the level tones, followed by four-year-olds with 31%, and then by six-year-olds. Adults, on the other hand, only produced 8% of level tones. These results are in line with previous findings about the prevalence of level tones in early child speech (Crystal 1979, Marcos 1987).

In the next section, we analyse quantitatively the scaling and then the alignment of the remaining 395 tokens that were produced with the falling tone.

5.1 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 vertical axis), the panel shows the F0 difference (in st) between H* and L% (on the horizontal axis) for English, Spanish and Catalan*
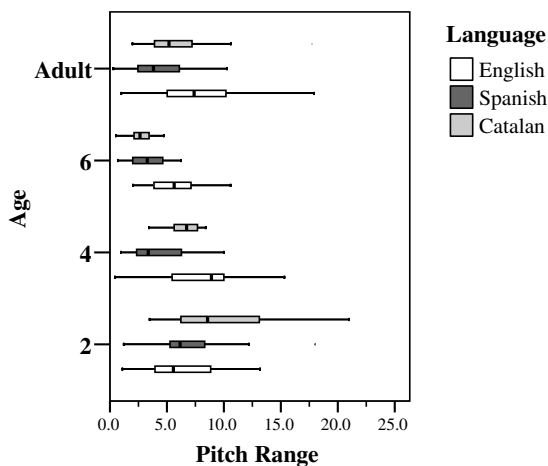
First of all, we notice from Figure 5 that the adult target is much higher in English (7.8 st) than it is in Catalan (6 st) and Spanish (4.9 st). For Catalan and Spanish children, achieving the target pitch range implies narrowing their pitch range while English children are very much closer to their language target. We also notice that the pitch range of the young children is not always narrower than that of the older children, as was also the case with level tones. In fact, the narrower pitch range corresponds to the six-year-olds. On average, two-year-olds have a pitch range of 7.4 st, four-year-olds of 6.9 st, six-year-olds of 3.9 st. Adults, on the other hand have a pitch range of 6.7 st. The wider pitch range of adults in comparison to six-year-olds can be due to the fact that adults are using child directed speech (CDS) as they are addressing young children. CDS is characterized by wider tonal inflections and higher pitch, among other features (see a prosodic analysis of CDS in Payne et al, submitted). It seems safe to asume that the target in normal, adult-directed speech should be closer to that of six-year-olds than to the adults in our data-base.

An ANOVA with dependent variable Pitch range and with between-subjects factors Age and Language confirm statistically significant effects of Age (F(3, 339)=14.164, p=0.000, partial eta-squared=0.115), Language F(2, 339)=11.984, p=0.000, partial eta-squared=0.068), and a significant Age by Language interaction F(6, 339)=4.638, p=0.000, partial eta-squared=0.078).

In order to check whether the stress pattern of the word had an effect and on pitch range as reported on the literature (Snow 1995, 1998, 2005), we broke up the data according to the position of the stress (Oxytone and Non-Oxytone), to Age, and to Language. Figure 6 shows the pitch range for each age group and for each stress pattern.

Figure 6   *For each age group (on the vertical axis) the panel shows the F0 difference (in st) between H* and L% (on the horizontal axis) for Oxytones and Non-oxytones.*
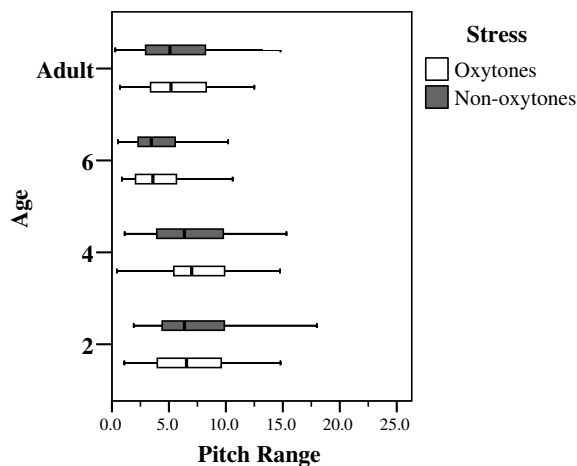
An ANOVA with dependent variable Pitch range and with between-subjects factors Age and Language confirm statistically significant effects of Age (F(3, 282)=11.009, p=0.000, partial eta-squared=0.113), Language F(2, 282)=98.664, p=0.001, partial eta-squared=0.056), and a significant Age by Language interaction F(6, 282)=3.399, p=0.003, partial eta-squared=0.073). Stress, on the other hand, is not significant and there are no significant interactions with Age or Language. The fact of whether the word is an oxytone or not, does not have significant effects on pitch range.

5.1   Alignment

As expected, the duration of the stressed syllable has an effect on the alignment of the high tonal targets. There is a significant correlation between the three measures of alignment and the duration of the syllable. Duration shows a strong negative correlation with H* to Syllable End (Pearson=-0.812, p<0.001), a medium negative correlation with H* to Syllable Onset (Pearson=-0.556, p<0.001) and significant though negligible correlation with L to Syllable Onset (Pearson=0.111, p < 0.05).

In order to check whether the presence of a sonorant onset and/or coda, and the stress pattern of the word had an effect and on alignment, possibly via the duration of the syllable, we conducted a MANCOVA (a multivariate ANOVA with one or more covariate factors), 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 Coda (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 ((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), L to Syll Onset (F(1, 299)= 7.752, p=0.006, partial eta-

squared=0.028). 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 significant though extremely small. Furthermore, it is counterintuitive that long syllables should show a more precise alignment of the L target to the onset of the syllable. We interpret this as a spurious effect of the measurement and segmentation criteria, as we segmented plosives from the burst, leaving out the closure interval. Syllables with a plosive in the onset are thus much shorter than syllables with a nasal or another sonorant.

The stress pattern of the word (oxytone, paroxytone, proparoxytone) has significant though modest effects on the alignment of the H* target only (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.

Finally, the type of coda does not have any significant effect on the three alignment measures. We expected that the amount of sonorant segmental material available would influence the alignment of the tonal target, as this effect has been reported in the literature for adult data (see Prieto and Torreira 2007). The type of coda has no significant effects on alignment, not even in the adult data only.

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. First we report the results for the alignment of the L target and then the results for the alignment of the H* target.

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.
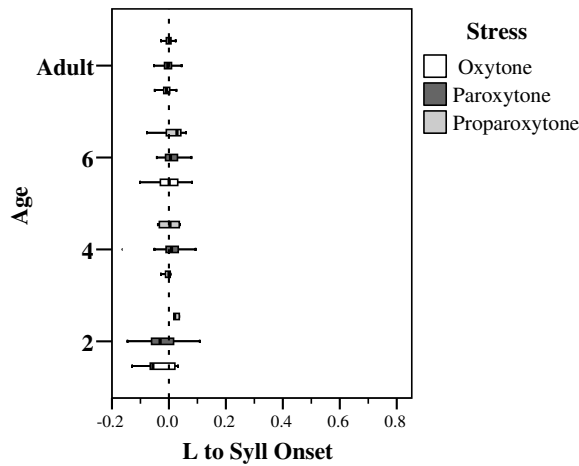
As we see in the boxplots in Figure 7 below, adults are much closer to the target than children are. At a glance, we appreciate as well cross-linguistic differences in both measures of alignment for L.

Figure 7   *For each age group (on the vertical axis,) the panel shows the distance (in seconds) from the L target to the syllable onset (on the horizontal axis) for English, Spanish and Catalan*
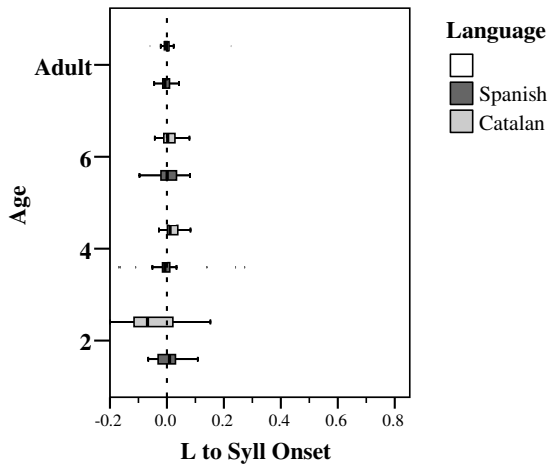
The analyses confirm statistically significant effects of Syllable Duration   (F(1,261)=16.504, p=0.000, partial eta-squared=0.061), Age F(3,262)=3.726, p=0.012, partial eta-squared=0.042), but not of Language, although the interaction of Age by Language is significant (F(3,262)=5.080, p=0.002, partial eta-squared=0.057). We ran Post hoc tests with Bonferroni correction to examine the interaction. In Catalan, 2-years olds show an average difference of 70 ms in their alignment of the L target, which is statistically significant. The other older children and adults show differences of about 4 ms, similar to the differences between the Spanish age groups, which are not significant. Figure 8 shows the alignment of L to Syll Onset for each age group and for each stress pattern.

Figure 8   *For each age group (on the vertical axis,) the panel shows the distance (in seconds) from the L target to the syllable onset (on the horizontal axis) for oxytones, paroxytones and proparoxytones*

All children show an alignment of the L target that is very close to the adult target. The analyses confirm statistically significant effects of Syllable Duration (F(1, 221)=13.602, p=0.000, partial eta-squared=0.061), and Stress F(2, 221)=3.224, p=0.000, partial eta-squared=0.030), but no significant effects of Age and no Age by Stress interaction.

As for the alignment of the H* target, we see in the boxplots in Figure 9 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 9   *For each age group (on the vertical axis) and each language (shaded boxes), the panel on the left shows the distance (in seconds) from the H* target to the syllable onset (on the horizontal axis) and the panel on the right the distance from H* to the syllable end.*
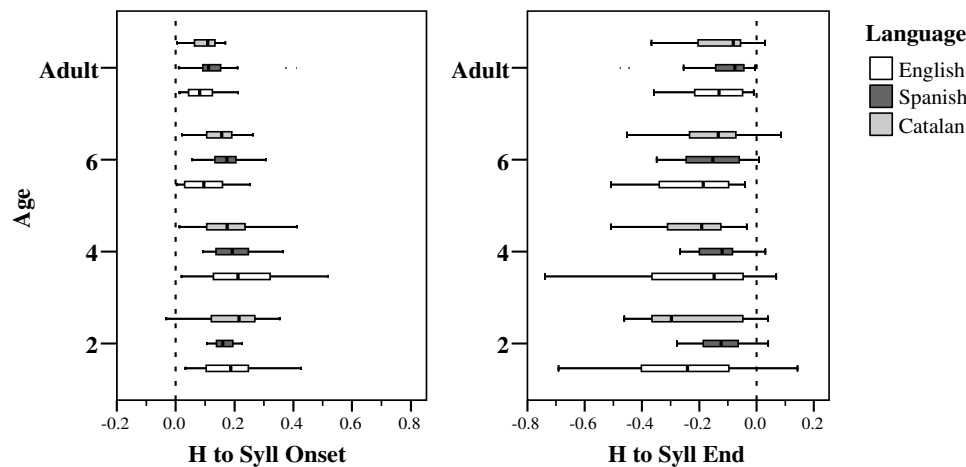

We notice as well much more variability, especially for the younger children, in the alignment of H* to Syll End (on the righ panel) than on the alignment of H* to Syll Onset (on the left panel). This is because children lengthen syllables much more than adults do and this in turn affects alignment. For this reason, we have included syllable duration as a covariate factor in the ANOVA analyses.

We thus ran a series of ANCOVAs for each H* alignment measure with between-subjects factors Age and Language and with covariate factor 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, 367)= 68.486, p=0.000, partial eta-squared=0.162), Age H to Syll Onset F(3, 367)= 10.642, p=0.000, partial eta-squared=0.083), and Language F(2, 367)= 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 significant effects of Duration (F(1, 367)= 864.775, p=0.000, partial eta-squared=0.709), Age (F(3, 367)= 8.488, p=0.000, partial eta-squared=0.067), Language (F(2, 367)=4.172, p=0.016, partial eta-squared=0.023), and of the Age by Language interaction (F(6, 367)=2.503, p=0.022, partial eta-squared=0.041). 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.

However, and except for Duration, we have found so far significant though very small effects of the independent variables on the alignment of H* to Syll End. This can be because we are throwing together words with different stress patterns. As the peak is aligned after the stressed syllable, the number of poststressed syllables will have an effect on its alignment. 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   *For each age group (on the vertical axis) the panels shows the distance (in seconds) from the H* to the syllable onset (on the horizontal axis) for oxytone, paroxytone and proparoxytone words and the duration of the syllable (in seconds, on the horizontal axis). 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.*
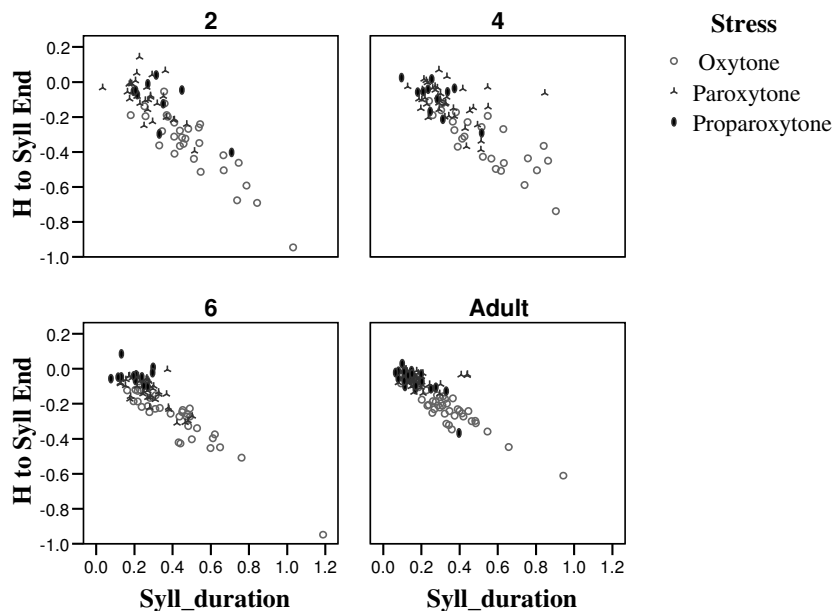
As we see in the panels in Figure 10, the stress pattern of the word has a rather large effect on the alignment of H* to Syll End (but not on the alignment of H* to Syll Onset, as we discuss below) and this effect is larger in older children and in adults. In oxytone words, the peak is retracted while on paroxytones and proparoxytones, the peak is aligned near the end of the syllable, but this 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 than even the younger children align their targets correctly near the end of the stressed syllable; they have already acquired adult-like patterns of alignment. Their problem is that they produce word-final syllables that are still too long and this excessive syllabic duration influences in turn the alignment of the H* target in relation to the end of the syllable.

This is confirmed by two ANCOVAs with dependent variables H* to Syll Onset and H* to Syll End, with between-subjects factors Age and Language 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, 325)= 122.288, p=0.000, partial eta-squared=0.281), Age F(3, 367)=4.457, p=0.004, partial eta-squared=0.041), and Stress F(2, 367)= 35.360, p=0.000, partial eta-squared=0.184). The interaction of Age by Stress is also significant (F(6, 325)=2.389, p=0.028, partial eta-squared=0.044). As for the alignment of H to Syll End, Syllable, we find significant effects for Duration (F(1, 325)= 449.004, p=0.000, partial eta-squared=0.589), Stress F(2, 325)=45.441, p=0.000, partial eta-squared=0.225), Age (F(3,

325)=2.460, p=0.000, partial eta-squared=0.023), and for the Age by Stress interaction (F(1, 325)=2.851, p=0.010, partial eta-squared=0.052). 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 is about 50 ms. The latter group also show less variability.

6 CONCLUSION

Using 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 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 produced a very large proportion of level tones, a variant of the falling accent with an extremely narrow pitch range. Other observations in the literature were not borne out by our data. Previous findings in English (Snow 1995, 1998) report that four-year-old children used a higher pitch range for paroxytones ("bottle") than for oxytones ("shoe") while adults used a similar pitch range for both. In our data, we found significant effects of Age and an Age by Language interaction. While English children are close to their language target, for Catalan and Spanish children, achieving the adult target implies narrowing their pitch range. We have also noticed that the pitch range of the young children is not always narrower than that of the older children, as six-year-olds have the narrowest pith range of all age groups. We attribute this to the fact that the adults are using a child-directed speech style, which is generally characterised, among other features, by wider pitch excursions. We assume, thus, that six-years old are closer to the normal (i.e., adult-directed) target.

Children also show excessive lengthening of final syllables. Despite this, children evidence 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 be realized both tonal targets in the one syllable available, is in itself remarkable.

Finally, our initial assumption that the prosodic typology of the L1 language would influence the acquisition of intonation, has been borne out by the data. Both in the analysis of pitch range and of the alignment of the H* target, we have found significant Age by Language interactions. Young Spanish children show a more precise command of pitch scaling and of the alignment of the high targets than Catalan and English children.

REFERENCES

Allen, G.D. and S. Hawking (1978) 'The development of phonological rhythm'. In A. Bell and J.B. Hooper (eds) *Syllables and Segments*. Amsterdam: North-Holland Publishing.

Archibald, J. (1995) 'The acquisition of stress'. In J. Archibald (ed.) *Phonological acquisition and phonological theory*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 81-109.

Arvaniti, A., Ladd, D.R., and I. Mennen (1998). 'Stability of tonal alignment: the case of Greek prenuclear accents'. *Journal of Phonetics* 26 .1:3-25.

Atterer, M. and D.R. Ladd (2004). 'On the phonetics and phonology of 'segmental anchoring" of F0: evidence from German'. *Journal of Phonetics 32.2:177-197.*

Beckman, M. and J. Pierrehumbert (1986) 'Intonational Structure in Japanese and English'. *Phonology Yearbook* 3: 15-70.

Boersma, P. and D., Weenink (2004) *Praat: doing phonetics by computer (Version 4.2.31)* {Computer program}. <http://www.praat.org/> (Retrieved April 30).

Chen. A. and P. Fikkert (2007) 'Intonation of early two-word utterances in Dutch'. *Proceedings of ICPhS 2007, 315-320.*

Crystal, D. (1978) 'The analysis of intonation in young children'. In F.D. Minifie and L.L. Lloyd (Eds) *Communication and cognitive abilities - early behavioral assessment*. Baltimore: University Park Pres, pp.257-71

Crystal, D. (1970) 'Prosodic systems and language acquisition'. In P.R. Léon, G. Faure and A. Rigault (eds), *Prosodic feature analysis* (Montreal: Didier), 77-90.

Crystal, D. (1979) 'Prosodic development'. In P.J. Fletcher and M.A. Garman (Eds), *Language acquisition*. Cambridge: CUP, pp.174-97.

Demuth, K. (1996) 'The prosodic structure of early words' In J. Morgan and K. Demuth *From signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp.171-84.

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

Estebas Vilaplana, E.; Prieto-Vives, P. **(**2008**).** 'La notación prosódica del español: una revisión del Sp_ToBI'. *Estudios de Fonética Experimental* XVII:263-283.

Fikkert, P. (1994) *On the acquisition of prosodic structure*. Dordrecht: Holland Institute of Generative Linguistics.

Frota, S., and M. Vigário (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.

Gerken, L. (1994) 'A metrical template account of children's weak syllable omissions from multisyllabic words'. *Journal of Child Language* 21:565-84.

Gerken, L. (1996) 'Prosodic structure in young children's language production'. *Language* 72:683-712.

Hallé, P. and 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.

Gili-Fivela, B. and M. Savino (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*. Causal Productions: Barcelona, pp.2933–2936.

Grabe, E. (1998) *Intonational phonology: English and German*. Doctoral dissertation. Max-Planck-Institut for Psycholinguistics and University of Nijmegen.

Grabe, E. and Post, B. (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*. Oxford University Press, pp.118-145.

Hochberg, J. (1987) 'The acquisition of word stress rules in Spanish'. *Papers and Reports on Child Language Development* 26:56-63.

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

Kehoe, M.M. and C. Stoel-Gammon (1997) 'Truncation patterns in English-speaking children's word productions'. *Journal of Speech, Language, and Hearing Research* 40:526-541.

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

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

Knight, R.A and F. Nolan (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., Faulkner, D., Faulkner, H. and A. Schepman (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., Meneken, I. and A. Schepman (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. (2002) 'The role of markedness in the acquisition of complex prosodc structures by German-Spanish bilinguals'. *International Journal of Bilingualism* 6:291–313.

Lleó, C. and Demuth, K. (1999) 'Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages'. In A. Greenhill, H. Littlefield and C. Tano (Eds.) *Proceedings of the 23rd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp.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. (submitted). 'Measuring rhythm in child-mother interaction". *Language and Speech.*

Pierrehumbert, J. (1980) *The phonetics and phonology of English intonation*. Tesis doctoral. MIT.

Prieto, P. and M. Vanrell (2007) 'Early intonational development in Catalan'. *Proceedings of ICPhS XVI,* Saarbrücken, Germany, 6-10 August 2007. Trouvain, J. and W. J. Barry (Eds.) Pirrot GmbH: Dudweiler, pp. 309-314.

Prieto, P. and F. Torreira (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 Vives, P., Aguilar Cuevas, L., Mascaró i Pons, I., Torres-Tamarit, F., and Vanrell, M.M. (2009) 'L'etiquetage prosòdic Cat_ToBI'. *Estudios de Fonética Experimental* 18: 287-309.

Estebas Vilaplana, E.; Prieto, P. (2009). 'La notación prosódica en español. Una revisión del Sp_ToBI'. *Estudios de Fonética Experimental* 18: 263-283.

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.

Ramus, F., Nespor, M. and Mehler, J. (1999) 'Correlates of linguistic rhythm in the speech signal'. *Cognition* 73.3:265-292.

Rietveld and Gussenhoven 1995 'Aligning pitch targets in speech synthesis: effects of syllable structure'. *Journal of Phonetics* 23.4:375-385

Schepman, A., Lickley, R. and 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., and J. Hirschberg (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. and H. Balog (2002) 'Do children produce the melody before the words? A review of developmental intonation research'. *Lingua* 112:1025-1058.

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 tones? '*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.

Santen, Jan P. H. and J. Hirschberg (1994) 'Segmental effects on timing and height of pitch contours'. *Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP-199)*, pp. 719-722.

Vihman, M. M., Nakai, S. and R. De Paolis (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.* Mouton de Gruyter: New York pp. 341-366.

Welby, P. and Loevenbruck, H. (2005) 'Segmental ''anchorage'' and the French late rise'. In: *Proceedings of Interspeech.* pp. 2369–2372.

Wells, B., Peppé, S. and A. Goulandris (2004) 'Intonation development from five to thirteen'. *Journal of Child Language* 31:749-778.

Xu, Y. (1998) 'Consistency of tone-syllable alignment across different syllable structures and speaking rates'. *Phonetica* 55: 179–203.

Xu, Y. and Liu, F. (2006) 'Tonal alignment, syllable structure and coarticulation: toward an integrated model'. *Italian Journal of Linguistics* 18.1: 125–159.

Corresponding author
Lluïsa Astruc

Spanish and Portuguese Department
Sidgwick Avenue
University of Cambridge
Cambridge
CB3 9DA
United Kingdom

mla28@cam.ac.uk

APPENDIX

Stress pattern	English	Spanish	Catalan
S	Sun Train Bee Moon Tea Knee Key	Sol Tren Pie Bus Sí Be Pe	Sol Tren Peu Bus Sí Be Pe
SW	Angel Baby Lyon Monkey Potty Daddy Mummy Money	Nena Luna Mono Llave Papi Mami Globo Bibe	Nena Lluna Mono Taxi Papa Mama Globus Bibe
WS	Guitar Balloon Giraffe	Bebé Camión León	Bebè Camió Lleó
WSW	Potato Banana Tomato Pajamas	Zapato Pelota Gallina Pijama	Sabata Pilota Gallina Pijama
SWW	Elefant Crocodile Melanie Butterfly	Áliga Música Mélanie	Àliga Música Mèlanie
WWS	Cockatoo Kangaroo Violin	Caracol Pantalón Violín	Cocodril Pantaló Elefant Ratolí
SWSW	Caterpillar Watermelon Helicopter	Mariposa Elefante Cocodrilo	Papallona Helicòpter Hipopòtam
SWSWW	Hipopotamus	Hipopótamo Helicóptero	