EARLY PERCEPTION OF INTONATION IN CATALAN AND SPANISH INFANTS

Eduard Artés Cuenca

Supervisor:
Dr. Pilar Prieto Vives

Research paper
MA in Cognitive Science and Linguistics
Universitat Pompeu Fabra

September 2010
# TABLE OF CONTENTS

1. INTRODUCTION ................................................................................................................ 3

2. EXPERIMENT 1 ................................................................................................................. 9
   2.1 GOALS .................................................................................................................. 9
   2.2 METHOD ............................................................................................................. 9
      2.2.1 PARTICIPANTS .................................................................................. 9
      2.2.2 MATERIALS ....................................................................................... 10
      2.2.3 PROCEDURE ..................................................................................... 12
      2.2.4 RESULTS AND DISCUSSION ......................................................... 13

3. EXPERIMENT 2 ............................................................................................................... 16
   3.1 PARTICIPANTS .............................................................................................. 16
   3.2 MATERIALS ................................................................................................. 17
      3.2.1 AUDITORY STIMULI ......................................................................... 17
      3.2.2 VISUAL STIMULI ............................................................................. 20
   3.3 EXPERIMENTAL CONDITIONS ........................................................................ 22
   3.4 OFF-LINE CODING ......................................................................................... 22
   3.5 PROCEDURE .................................................................................................... 22
   3.6 RESULTS ........................................................................................................ 24
   3.7 DISCUSSION .................................................................................................... 27

4. CONCLUSION ................................................................................................................. 29

5. ACKNOWLEDGEMENTS ............................................................................................... 31

6. REFERENCES ............................................................................................................... 32
1.- INTRODUCTION

One issue which has attracted attention in research on language acquisition is the early perception of stress. Early discrimination studies on the perception of stress have found that newborns can discriminate between trochaic and iambic patterns. Sansavini, Bertoncini & Giovannelli (1997) reported the ability of Italian newborns to discriminate stress patterns in different pseudo-words, regardless of consonantal variation in the speech materials. Infants between one and four months can detect differences in stress location in disyllabic pseudo-words (Jusczyk & Thompson, 1978), even with synthesized speech (Spring & Dale, 1977). Similarly, Weber et al. (2004) examined with ERPs early sensitivity to trochaic and iambic stress patterns in two syllable pseudo-words with German adults and infants at the ages of 4 and 5 months. The Mismatch Negativity (MMN), a response in the brain to a deviant stimulus in a series of standard stimuli, has been used to investigate some phonemic differences and other processes related to audition. The results showed that German adults were capable of discriminating between the two stress patterns, whereas 5-month-old infants only distinguished the trochaic deviant item (trochee condition). 4-month-olds, however, were incapable of distinguishing any deviant items at all. The authors offer two possible explanations: a) “the trochaic item consisting of a long syllable as its onset could be more salient and thus more easily discernible” (p. 160); or b) the trochaic stress pattern of German, the infants’ native language, is detected more easily because of prolonged prior exposure.

Though attunement of stress perception abilities relative to the native language begins at 4/5 months of age, during the second half of the first year of life (at around 9 months) infants are already aware of the stress properties of the words of their own language and show clear preference behaviours. Lexical stress preference studies have observed emergent attention biases in infants well before their first year of age. In one of the first preference studies, Jusczyk, Cutler & Redanz (1993) demonstrated that 9-month-old infants preferred strong-weak words (= trochaic), but 6-month-olds did not. They thus argued that preference is shaped by experience, indicating that the “trochaic bias” is not universal. Jusczyk, Houston & Newsome (1999) found that, at 7.5 months
of age, infants could only segment disyllabic words with a trochaic stress pattern, while iambic words were missegmented (the final stressed syllable was taken as the first one). This strategy for segmenting words (the so-called Metrical Segmentation Strategy proposed by Cutler & Norris, 1988) corresponded to the most frequent stress pattern of their native language (English), so there was a great likelihood of strong syllables marking the occurrence of a word boundary. Nevertheless, this tendency changed at 10.5 months of age, because the children had already acquired other strategies to segment iambic words correctly, thus indicating that they had begun to rely on other sources of information, such as allophonic or phonotactic cues. In an attempt to contrast these findings with data from languages other than English, Pons & Bosch (2007) ran an experiment to test the validity of the “stress cue” in syllable-timed languages such as Catalan and Spanish. Although there are some differences between Catalan and Spanish (Catalan has more iambs and monosyllabic words), both languages are usually considered basically trochaic. However, unlike English, they do not have a strong initial lexical stress pattern, so the authors hypothesized that the “stress cue” would be less reliable in Catalan and Spanish infants. They tested a group of 6-month-old and 9-month-old children using a slightly modified version of the Head Turn Preference (HTP) procedure. The results were conclusive: Catalan and Spanish children do not show any preference for Strong-Weak (SW) syllables at either 6 or at 9 months. However, when preference was explored with Spanish materials containing heavy syllables in stressed position (CVC.CV and CV.CVC, respectively), both a trochaic and an iambic preference could be elicited (Pons & Bosch 2010). These results suggest that Spanish children at 9 months of age have knowledge about the highly regular link between heavy syllables and stress assignment and this determines infants’ preferences at 9 months of age.

Most research on early perceptual development of prosody has been concerned with discriminability and preference patterns in relation to stress contrasts, and little is known about the early perception and acquisition of intonational patterns. The evidence thus far suggests that children can clearly discriminate between target intonational contrasts. Nazzi, Floccia & Bertoncini (1998) investigated whether French newborns were sensitive to the difference in pitch height and whether they could
extract this information from lists of Japanese words. 121 newborns were tested (using the High Amplitude Sucking -or HAS- paradigm) with two lists of 24 bisyllabic (bimoraic) Japanese words differing in pitch contour (H-L vs. L-H, where H stands for “high tone” and L for “low tone”) and the expected result was found: French newborns could indeed discriminate differences in pitch (they extracted the common pitch contour dimension of the words in each list). The authors “suggest that this sensitivity is part of the universal repertory of infants’ innate abilities” (p. 782). Sato, Sogabe & Mazuka (2006) investigated pitch contour perception with 4-, 6- and 8-month-old Japanese infants. Using the HTP procedure with tokens taken from Nazzi, Floccia & Bertoncini (1998), they tested Japanese infants with words differing in pitch accent. Surprisingly, 4-month-olds were capable of distinguishing the target word pitch patterns of H-L vs. L-H, and 6- and 8-month olds could even do it within varied lists of words. More recently, Soderstrom, Ko & Nezvorova (2010) investigated infants’ early perception of intonation through a task in which infants had to discriminate question from statement intonation. English infants were tested using a habituation paradigm with normal and low-pass filtered speech, and the results showed that they were able to discriminate statements from yes-no questions but, surprisingly, not from wh-questions. They attributed this result to the fact that yes-no questions display a more exaggerated pitch contour.

With respect to word learning, research has focused on infants’ ability to learn phonetically similar words, especially regarding consonants (Fennell, 2004; Werker et al., 2002; Pater, Stager & Werker, 2004) and vowels (Mani & Plunket, 2008). These investigations seemed to show that 14-month-old children were not capable of learning two new words (i.e., associating new words with target objects) that differed in a single phonological-feature change (e.g., Stager & Werker 1997, among many others). Yet recent evidence has shown that 14-month-olds are capable of learning and mapping two similar new target words to different objects when using a visual choice method rather than the standard switch task (e.g., Yoshida et al., 2009). To our knowledge, only one study has carried out a word-learning task with a lexical stress contrast (and using statement intonation). Curtin (2009) tested sixteen 14-month-old English infants with new word–object associations that differed only in stress pattern.
(Sww versus wSw; “BEdoka” versus “beDOka”\textsuperscript{1}). Results showed that, even without contextual support, infants at 1;2 succeeded at this task. Mean looking times were later analyzed, and confirmed the idea that infants can store stress properties at a very young age. The results of Curtin’s study suggest that the salient acoustic properties associated with lexical stress might be facilitating the task of word–object association in a stronger way than what is allowed by a vocalic or consonantal phonemic contrast.

The present study has the goal of further investigating whether children’s interpretations of phonological contrasts based on stress can take into account intonational variation present in the input language. In other words, we would like to investigate whether children at 14-18 months have learnt what we call the “intonational grammar” of their input language. As is well-known, in intonational languages different stress patterns (Sw WS, etc.) can be pronounced with different intonation patterns (rising, falling, rising-falling), and these intonational differences convey not contrasts in lexical meaning (as in tonal languages) but rather differences in phrasal and attitudinal meaning. If children have learnt that intonational variation does not affect lexical meaning they will be able to rightly associate several intonation patterns with their corresponding contrastive stress patterns (wS and Sw, for example). Two recent studies have yielded partially contradictory evidence in this respect.

In the first of these experiments, Quam & Swingley (2010) tested the role of pitch modification in a word-learning task with 24 adults and 48 2.5-year-olds using an eye-gaze procedure in which participants had to recognize a previously taught word that had been slightly modified. A language-guided looking procedure was used to assess adults’ and infants’ interpretation of phonological (vowel quality) versus non-phonological (pitch-contour) changes in a newly learned word (“deebo”). Participants listened to the trained word 8 times in the test trial and 5 times in the mispronunciation condition (“dahbo” for the phonological contrast and “deebo” with a different intonation contour from that learned in the labelling task—rise-fall contour or low fall contour—for the non-phonological contrasts). The time spent looking at

\textsuperscript{1} Capital letters throughout the article will indicate stressed syllables.
pictures shown on a screen were measured for both age groups, confirming the hypothesis: changes in pitch contour do not impair recognition (but changes in vowel quality do). These findings demonstrate that 2.5-year-old children do not store pitch and other kinds of phonetic information in their lexical representations if they are not used phonologically (pitch variation is not always used phonologically in intonational languages, as opposed to tonal languages).

The results of a second more recent experiment (Quam 2010) carried out with adults, 2.5- and 5-year-olds showed that children under age four have not fully developed the skills of pitch cues to lexical stress. Participants looked at different pictures of a bunny and a banana over sixteen trials, with matched audio cues differing only in stress, which was marked through pitch (BUunny vs. buNNY and BAanna vs. baNAna). The comparison between attention paid to the correct-stress items and the mispronounced-stressed items was measured using an eye-tracker. While the mean difference between correct pronunciation and mispronunciation was nearly 20% and 10% in adults and children over 4, respectively, in children under 4 it was only 2%. Thus, children over four seem to behave more or less like adults, appearing sensitive to mispronunciations in both conditions, but the mispronunciation effect in children under age four is clearly lower. The authors hypothesize that sensitivity to pitch variations might be acquired late due to its relative low validity in the assignment of stress (which is accompanied by vowel reduction, duration and so on). They argue that “age four is a developmentally turning point in the interpretation of pitch both for stress and for emotional prosody—in contrast to the early acquisition of consonant, vowel, and lexical-tone categories.”

The specific aim of the present study is to investigate whether infants as young as 18 months old are able to associate different intonation patterns with the same word in a word-object association task. Two lines of evidence suggest that children might be capable of taking into account intonational variation of their input language—in word-learning tasks. First, exaggeration and variation of pitch contours in the presentation of the experimental materials (and which is typical of Infant-Directed Speech or IDS, i.e., speech directed at infants by adults) is widely regarded as facilitating infants’ linguistic
learning. Thiessen, Hill & Saffran (2005) tested two groups of infants to determine the
effects of IDS and Adult-Directed Speech –ADS– in word segmentation. The infants had
to listen to words and part-words, and results showed that the infants were able to
detect word boundaries (i.e., showed a preference for words over part-words) after
exposure to IDS not after hearing ADS. IDS was thus shown to facilitate word
segmentation and, therefore, language acquisition.

Second, production studies of early intonational patterns have shown that infants as
young as 14 months can produce a handful of target intonational contours and by the
time they are 2 already master the production of the basic intonational grammar of
Prieto & Vanrell (2007) investigated intonational development in four Catalan-
speaking children between the ages of 12 and 26 months. Independent of their
acquisition of grammar (some children reached the two-word period earlier than
others), all infants demonstrated a wide repertoire of intonations. The most widely
used contour was the statement, but they could also produce vocatives, insistent
requests and insistent vocatives, as well as express linguistic focus.

The present study consisted of two experiments. The main goal of Experiment 1 was to
select the appropriate intonational contours to be used for Experiment 2, in which
infants would participate. Since there is some degree of ambiguity in the perception of
stress contrasts when several intonation contours are used, as will be explained below,
an important step in this direction was to test the degree of discriminability across
stress contrasts with several intonation patterns. For our study with infants, we
therefore intended to select target intonational contrasts that allow stress contrasts
to be successfully discriminated by adults and also a target contour that is easily
misinterpreted by adults as a control (as we expected that infants would also have
problems identifying stress contrasts in this intonational condition). For its part, as
noted above, the main goal of Experiment 2 was to determine the development of
intonational grammar in 18-month-olds. To that end we planned to test Spanish- and
Catalan-speaking infants using an associative task with a standard switch paradigm (in
a visual choice design). At this age, infants should be capable of discriminating similar
sounding words (differing only in stress) and associating certain common intonation contours with recently learned concepts. Infants’ looking times to correctly associated objects would then be recorded and analyzed.

2.- EXPERIMENT 1

2.1.- Goals: There was one important reason to test adults’ identification of stress patterns under different intonation conditions before testing children. Since there is some interaction between intonation and stress, in some cases trochaic stress patterns produced with complex intonation patterns cannot be easily discriminated, even by adults. In order to guarantee that the association task would be straightforward in the materials for the children’s experiments and avoid any ambiguities, we first designed a stress identification task for adults. The idea was that the results of this experiment would help us select the target intonation contours to be used in the children’s experiment.

Therefore, as noted, the main goal of this experiment was to investigate Catalan adults’ ability to identify stress contrasts (using SW and WS stress patterns) when pseudo-words were pronounced with the different intonation contours present in the intonational grammar of their native language.

2.2.- Method

2.2.1.- Participants: Twenty-two adults, ten males and twelve females, all native Catalan-speakers, participated in the identification experiment. In Catalonia, and especially in the Barcelona area, both Catalan and Spanish are widely spoken and both

---

2 It is important to note here that in real life situations the context can help the listener discriminate the target stress pattern, as different pragmatic meanings are associated with every intonation contour.
languages are taught at school. Though all of the participants were bilingual, they considered themselves Catalan-dominant, and Catalan was the only language spoken at their homes. The subjects belonged to different Catalan-speaking areas, covering the whole Catalan linguistic domain: 16 were speakers of Eastern Catalan (15 spoke the Central dialect and 1 the Balearic dialect) while 6 were speakers of Western Catalan (3 spoke the Tortosí dialect, 2 the North-Western dialect and 1 the Valencian dialect). Participants were between 20 and 58 years old, and all but two were postgraduate or undergraduate students.

2.2.2.- Materials: Two native Catalan-speakers (one male and one female) produced natural tokens of the pseudo-word “dibu” with seven different intonation patterns and in two stress conditions, trochee (DiBu)—that is, with stress on the first syllable—or iamb (diBU)—with stress on the second syllable. These intonation patterns were selected because they are commonly used intonation patterns in the adult language. The stimuli were recorded as Adult-Directed Speech albeit in an emphatic fashion using professional equipment in a recording studio.

Table 1 shows the schematic representation of the seven nuclear pitch contours with the trochee condition (Dibu) together with their Cat_ToBI labels and one of their most common pragmatic meanings. For the intonation analysis, we used the Catalan version of ToBI, Cat_ToBI (Prieto et al. 2009; Aguilar et al 2009; Prieto in press). The darker region in the schematic representation stands for the stressed syllable (which is symbolized by an asterisk “*” in the schematic ToBI representation), while the lighter one stands for the unstressed syllable (symbolized by a percentage sign “%”, which marks the boundary tone). The letters “L” (low) and “H” (high) indicate the different tones associated with each syllable, tones which can also be combined. Thus, L+H* indicates that a low tone is aligned with the start of the stressed syllable and rises to a high tone within the same syllable. On the other hand, L* HH% is phonetically realized

---

3 Boundary tones are distinguished from other tones by their distribution. Unlike stress-dependent intonational pitch accents, boundary tones only occur on the edgemost syllable of a domain, such as the last syllable of an intonational phrase. Unlike lexical tones, boundary tones have a phrasal distribution and function to distinguish sentence types (Myers, 2004:147).
as a low tone only on the stressed syllable, which rises to a high one during the posttonic syllable.

<table>
<thead>
<tr>
<th>Request Type</th>
<th>Cat_ToBI Label</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insistent request</td>
<td>L+H* LHL%</td>
<td><img src="insistent_request.png" alt="Insistent Request" /></td>
</tr>
<tr>
<td>Gentle request</td>
<td>L* HL%</td>
<td><img src="gentle_request.png" alt="Gentle Request" /></td>
</tr>
<tr>
<td>Vocative</td>
<td>L+H* M%</td>
<td><img src="vocative.png" alt="Vocative" /></td>
</tr>
<tr>
<td>Information-seeking yes-no question</td>
<td>L* HH%</td>
<td><img src="yes-no_question.png" alt="Information-seeking yes-no question" /></td>
</tr>
<tr>
<td>Invitation yes-no question</td>
<td>L+H* HH%</td>
<td><img src="invitation.png" alt="Invitation yes-no question" /></td>
</tr>
<tr>
<td>Narrow focus statement</td>
<td>L+H* L%</td>
<td><img src="narrow_focus.png" alt="Narrow focus statement" /></td>
</tr>
<tr>
<td>Echo question (counterexpectational meaning)</td>
<td>L+H* LH%</td>
<td><img src="echo_question.png" alt="Echo question" /></td>
</tr>
</tbody>
</table>

**Table 1.** The table contains the schematic representation of seven Catalan nuclear pitch contours included in Experiment 1, their Cat_ToBI labels and one of their common pragmatic meanings.

Figure 1 exemplifies four of the contours present in the target materials. The information-seeking yes-no question (top left panel), is characterized by a low tone associated with the stressed syllable, with an extra high boundary tone on the word edge. The insistent request (top right panel) has a rising movement from low to high on the stressed syllable plus a complex rising-falling movement –LHL- as boundary tone. The gentle request (bottom right panel) is characterized by a rising movement on
the stressed syllable and on the boundary tone. Finally, the echo question (bottom left panel) starts low on the stressed syllable and then has a falling contour –HL- as boundary tone.

Figure 1 Waveform, f0 contour and spectrogram of four intonation patterns present in the target materials. Insistent request (L+H* LHL%, top right panel), echo question (L+H* LH%, bottom left panel), yes-no question (L* HH% top left panel) and gentle request (L* HL%, bottom right panel)

2.2.3.- Procedure: Audio stimuli were set up using E-prime. The experiment consisted of 4 blocks, the first one being the familiarization phase, which were separated by 10-second rest periods. In each block the seven stimuli were repeated twice in each condition, that is, 7 target intonations (see the Materials section) x 2 stress conditions x 2 different readers. Altogether, a total of 112 semi-randomized stimuli were presented to each listener.

Participants sat in front of a computer, in a quiet room, with headphones. They were asked to indicate whether they heard the word as a trochee (SW) or as an iamb (WS)
by pressing a different keyboard button (P / F) for each response. The “P” key stood for the trochaic condition (Principi in Catalan – beginning) while “F” stood for the iambic condition (Final in Catalan – end). The experiment lasted approximately 15 minutes.

2.2.4.- Results and discussion: A factorial ANOVA was applied to the data. The dependent or response variable was the rate of identification of stress (trochee = 1; iamb = 2) and the main factors were stress pattern (2 levels: trochee and iamb), and intonation (7 levels: insistent request, gentle request, vocative, yes-no question, invitation yes-no question, narrow focus statement and echo question). The analysis revealed significant main effects of the stress pattern of the stimulus word on stress identification (F(1, 216.284)=2370.994, p<.0001) and of the intonation pattern on stress identification (F(6, 6.599)=72.339, p<.0001). This analysis also showed that there was a significant interaction between the stress and intonation factors (F(6, 4.121)=4.920, p<.0001). Post-hoc analyses revealed that the behaviour of two intonation patterns, namely those of the gentle request and insistent request were significantly different from the rest of the intonation contours (at p<.05; see also Figure 1). The statement intonation was only significantly different from the yes-no question (p<.01), and the invitation yes-no question differed significantly only from the vocative (p<.0001). The yes-no question, in turn, was also significantly different from the echo question (p<.01).

Figure 2 shows the mean stress responses (trochee = 1; iamb = 2) as a function of the seven intonation conditions (x-axis). Basically, adults discriminated between the two stress conditions (DiBu vs. diBU) correctly in the majority of the intonation conditions. As the graph shows, participants showed a clear stress classification problem with the gentle request intonation (L* HL%): in this case, the target word DiBu with a L* HL% intonation was systematically interpreted as a word with iambic stress (WS). Insistent requests, invitation yes-no questions and echo questions also triggered some degree of confusion in the stress classification task, even though the stress difference was also clear. Finally, stimulus words with statement, yes-no question or vocative intonations caused no problems in the stress classification task. For this reason, they were chosen
as the stimulus intonation patterns for the subsequent experiment involving infants (Experiment 2).

Why do some of the stimulus intonation patterns cause a degree of confusion in the stress classification task? Our results show that the trochees are more easily confused than iambs. In the iambic patterns, pitch movements (pitch accents and boundary tones) are concentrated on the second and last syllable and thus the salience is generally perceived in this final syllable. By contrast, in the trochee condition when the boundary tones are complex they generally appear in the post-tonic syllable, and this complex configuration can presumably misguide the perception of stress.

For example, let us compare the trochaic target word DIbu with gentle request intonation (L* HL%; left panel, Figure 3) with the iambic word diBU with statement intonation (L+H* L%; right panel, Figure 3). We argue that the rising movement in the post-tonic syllable in the gentle request intonation (left panel) is the key feature

![Figure 2. Mean stress identifications (trochee = 1; iamb = 2) as a function of the seven intonation conditions (x-axis) and the two stress conditions (trochee, iamb).]
leading to the reinterpretation of this stimulus as an iambic stressed word with statement intonation (right panel).

Figure 3. Waveform, spectrogram, and f0 contours of the gentle request intonation in the trochaic word DiBu (left panel) and of the statement intonation in the iambic word diBU (right panel.).

The intonation patterns that caused some degree of confusion in the stress classification task (i.e., insistent requests, invitation yes-no questions and echo questions) share the property of having complex boundary tonal movements on the final syllable, which can induce the perception of stress in this final syllable and thus a reinterpretation of trochees as iambs. It is worth pointing out, however, that this ambiguity phenomenon probably does not reflect a common situation in online speech processing because minimal pairs of stress contrasts are relatively scarce. In spontaneous situations cues such as facial gestures or discourse context would help the listener interpret the target intonation contour, but when heard in isolation, as was the case in our experimental set-up, the rising boundary tone (LH%) of a trochaic echo question (see figure 4), for example, can induce misinterpretation of stress, making the unstressed syllable more prominent and thus easily confused with an iambic yes-no question (L* HH%)
In sum, the statement, vocative and yes-no question intonation contours were the most successful ones in the stress discrimination task by adults and were therefore chosen for our subsequent child experimental task.

3.- EXPERIMENT 2

In this experiment five infants from both Catalan- and Spanish-speaking home environments were tested on their capacity to associate different intonations with a newly learned word. Our goal was to determine whether 18-month-olds had already acquired the intonational grammar of their native language, in contrast with other findings which have determined that intonational development is not stable until age four (Quam 2010).

We chose a visual choice design in this experiment to more easily detect the infants’ preference for the correct object. The Spanish adaptation of the MacArthur Communicative Development Inventory (López-Ornat et al., 2003) was used to assess infants’ prior lexical knowledge. We also attempted to record infants’ speech production before and after the perceptual experiment, in order to assess their word inventory. However, this attempt was not very fruitful due to the difficulty of making infants speak in a novel environment.

---

Figure 4. Waveform, spectrogram and f0 contours of the trochaic echo question.

---

4 This was our original intention. As will be seen, the age range was widened to include infants aged 14 to 19 months due to the limited amount of time available to us.
3.1.- **Participants:** Participants were recruited by visiting kindergartens in the Barcelona area, with parental consent acquired prior to the experiment. Nine infants were originally selected, but in the end four were insufficiently at ease in the experimental context to participate. Three of the five infants who did participate came from a Catalan-speaking home environment, one came from a Spanish-speaking environment and the fourth was being raised in a bilingual Catalan-Spanish environment. All participants were between 14 and 19.23 months old.

3.2.- **Materials**

3.2.1.- **Auditory stimuli:** A Catalan-speaking adult female produced the two stimulus words (DIbu vs. diBU) with the following intonation patterns, using Infant-Directed Speech:

- a) statement (L+H* L%)
- b) vocative (L+H* M%)
- c) information-seeking yes-no question (L* HH%)
- d) insistent request (L+H LHL%)

As mentioned above, the choice of pitch contours was based on the results of Experiment 1, i.e., we selected the three that had achieved the best stress discrimination scores among adults (see figure 5), namely, the information-seeking yes-no question (L* HH%), vocative (L+H* M%) and statement (L+H* L%). In addition, the insistent request was included precisely due to the confusion it elicited in the adult experiment: we hypothesized that if infants have already acquired the intonational
grammar of Catalan and/or Spanish, they should have the same problems in discriminating the stress of the insistent request contour as adults.

The final audio track consisted of 5 blocks, as follows:

a) **Block 1**. The pseudo-words Dibu and diBU were contextualized by using the following three utterances. Each utterance was repeated three times:

- *Mira, un Dibu / Mira, un diBU* - Look, a Dibu / Look, a diBU
- *És un Dibu / És un diBU* - It’s a Dibu / It’s a diBU
- *Dibu! / diBU!* - Dibu! / diBU!

**Figure 5.** Waveform, spectrogram, and f0 contours of the yes-no question in the trochaic word Dibu (top left panel), the trochaic statement (top right panel) and trochaic vocative (bottom). The differences in HH% (yes-no question), L% (statement) and M% (vocative) boundary tones can be seen clearly. Presumably, these distinct tonal heights make discrimination easier for hearers.
In the first two utterances the intonation contour of DiBu /diBU corresponded to a statement (L+H* L%), while the last one was that of a vocative (L+H* M%).

b) **Block 2.** The pseudo-words DiBu and diBU were again contextualized in three different utterances, but with certain variations. Each utterance was again repeated three times:

- És un DiBu? / És un diBU? - Is it a DiBu? / Is it a diBU?
- DiBu? / diBU? - DiBu? / diBU?

In all three cases here, the intonation contour of DiBu /diBU corresponded to a neutral yes-no question (L* HH%).

c) **Block 3.** The pseudo-words DiBu and diBU were played in isolation. The intonation contour corresponded to an insistent request (L+H* LHL%). The isolated word was repeated five times.

d) **Block 4.** This utterance, the word “hola” (hello), was played in isolation. This block was used as a pre- and post-test stimulus.

e) **Block 5.** This block was intended as a filler to maintain the infants’ attention between test segments. It was therefore matched with an image depicting two familiar objects, a car and a pair of shoes. The infants heard a voice referring to the car and it was assumed that they would direct their gaze to the corresponding picture. The utterance was repeated three times:

- I el cotxe? - And the car?
- Mira el cotxe! - Look at the car

All the auditory stimuli were recorded professionally in a studio and lasted 20 seconds each. In each block the various stimuli were preceded, followed and separated by
pauses of slightly differing durations. In Block 1 there were initial and final silences of 0.75 seconds, a 1-second pause between one set of three sentences and the next and also a 1-second pause between the sentences in each set. Block 2 had a 1.5-second silence at the beginning and end, a 2.1-second inter-set pause and a 1 second pause between each of the three sentences. In Blocks 3 and 4 the initial and final silences lasted 1.5 seconds while the inter-set pause lasted 3.1 seconds. Finally, in the filler block we put a 1.8-second silence at the beginning and end, a 3-second inter-set pause and a pause of 1.3 seconds between the sentences of each set.

3.2.2.- **Visual stimuli:** During the test, the audio materials were accompanied by the projection of either a static image or an animated movie. The static visual materials showed images of two imaginary objects which children were expected to pair up respectively with the audio labels Dlbu or diBU. There were four different animated movies: two habituation videos, one pre- and post-test video and one “attention-getter”. Like the audio stimuli, each visual stimulus lasted 20 seconds.

For the habituation segment, an animated video movie was created using the two geometrical objects seen in figure 6. In the movie, one or the other of these two objects moved slowly but constantly back and forth across a black background while the “name” of the object—either “Dlbu” or “diBU”—was simultaneously played on the audio track [= Block 1].

After this habituation session, an image was projected showing simultaneously the two objects seen in the training phase, this time stationary rather than moving, with the “Dlbu” object to the left and “diBU” on the right, or vice versa (see figure 8). While this image was projected, audio Blocks 1 or 2 were played back. It was expected that the infant would direct his/her gaze to the object named in the audio track, i.e., to either the right- or left-hand side of the screen.

The filler movie had the same characteristics as the habituation video but, as noted, the images depicted familiar objects (a pair of shoes and a car - see Figure 9).
Simultaneously, audio Block 5 was played back during this segment of the experiment. Images and videos covered the whole monitor size.

In the pre- and post-test movies, the object shown in figure 7 moved vertically up and down the screen while the audio track consisted of Block 4, “hola” (hello).

Figure 6. Objects shown during habituation and test segments of Experiment 2: “Dibu” and “diBU”

Figure 7. Object shown during pre- and post-test segments of Experiment 2

Figure 8. Stationary images used during test segments of Experiment 2

Figure 9. Stationary images used during “attention-holder” segments of Experiment 2
3.3. **Experimental conditions:** The study was conducted at the *Laboratori de Recerca Infantil* at the *Universitat Pompeu Fabra (Campus de la Comunicació-Poblenou)* in Barcelona, in a small room with two 20W lamps illuminating each side of the room, creating a semi-dark atmosphere to minimize distraction on the part of participants. Images were shown on the big screen monitor of a Macintosh computer, while a digital video camera unobtrusively mounted above the screen recorded the infant’s face through a small hole in the grey cloth partition that formed a square but unroofed space constituting about ¾ of the studio. Two stereo loudspeakers hidden behind the partition played the stimuli at 65 (+-5) dB SPL. During the experiment, infants sat on their parent’s lap at a distance of 1 metre from the screen, and the images of the child’s face recorded by the camera were sent to a control computer. The experiment was run using Habit Software (Cohen, Atkinson & Chaput, 2004).

3.4. **Off-line coding:** Infants’ looking times were coded offline by a coder who was unaware of the nature of the experiment and the stimuli presented to the child. The playback stimuli were matched with the eye movement recordings once the looking times had been annotated. Matched recordings were coded frame by frame (1 frame = 40 ms).

3.5. **Procedure:** We used a visual habituation procedure, one of the most common techniques in infant behavioural studies. We implemented a switch procedure (Werker et al., 2002) with some modifications involving a visual choice design (Yoshida et al., 2009). Looking times were calculated over a window of 4 trials and the habituation criterion was set at 65% of decrement.

The experiment began when infants visually fixated on a yellow dot flashing on the monitor (also used as an attention-getter between trials). In the meantime, the experimenter (out of sight behind the partition began to monitor infants’ looking times and attention through the keyboard and then initiated playback of the pre-test video (the object seen in figure 7 with “hola” played on the audio track). This was followed by the habituation trials, during which time infants were trained to match the two pseudo-words DiBu and dBiU with the auditory stimuli of block 1 (see materials
section). However, half of the infants were trained to match DIbu with the round-edged figure and diBU with the square-edged one while the remaining infants learned to match DIbu with the square-edged figure and diBU with the round-edged one. If the infant looked at a projected figure for less than 1 second during any habituation trial, the trial was repeated. A maximum number of 24 habituation trials was allowed.

The test phase consisted of 10 choice design trials preceded by 1 filler (car and shoes images). Four of the test trials were “same” and the other four were in the “switch” condition (with the auditory stimuli of block 2), all randomly ordered. The last two trials were also in the switch condition but with the isolated ambiguous intonations of block 3 on the audio track.

For each child, before carrying out the perception test, we conducted a short production session in order to informally evaluate the children’s production of lexical and intonation patterns and possibly relate their perception results to their production patterns. Following written instructions, parents were asked to recreate four different situations meant to induce the child to produce four intonation contours (a question, statement, insistent request and vocative) in a 5-minute recording session. The instructions given to parents were as follows:

a) Object identification: show this toy to your child and try to make him say its name (statement).

b) Take the toy away from your child and try to make him/her ask for it (insistent request).

c) Try to make your child call a distant object (vocative).

d) Hide a toy and try to make your child ask where it is (question).

This attempt to induce vocalizations proved to be difficult because the infants were in a new environment, which inhibited them. Although parents tried hard to make them
produce utterances, little information about intonation contours could be extracted. Utterances produced by the respective infants were as follows:

- **Àlex**: 2 questions and 5 statements (nonsense words)
- **Armand**: 6 statements and 1 vocative (Catalan words), 1 insistent request (nonsense word)
- **Laura**: 2 statements (Spanish words)
- **Emma**: 6 statements (Catalan words) and 1 surprise intonation contour (Catalan nonsense word)
- **Iker**: 4 statements (nonsense words)

Although the parents of some of the infants reported that their children spoke a great deal at home, the Mean Length of Utterance (MLU) was 1 word/morpheme, basically because of the inhibitory effect of the experimental context.

The MacArthur inventory was also given to parents, who were asked to return the booklets to the university after their completion.

### 3.6.- Results

To check the infants’ attention throughout the experiment we compared looking times during the test trials with looking times during the post-test\(^5\). Mean looking times as measured in seconds were 8.848 s in the test trials and 18.036 s in the post-test, thus showing that infants were indeed involved in the task.

The main data analyses consisted of comparing the differences in infants’ looking times between correct (i.e., matched audio name and visual figure) and incorrect (i.e., mismatched name and figure) items during test trials. In none of the three intonation conditions was a significant difference found (statement/vocative: t(4)= 1.221, p= .289, Mean\(_{\text{correct}}\): 5.738 - Mean\(_{\text{incorrect}}\): 4.196; yes-no question: t(4)= 1.029, p= .361,\

\(^5\) Since we did not measure the novelty response of infants (we did not use the original switch paradigm), we did not run a comparison between the pretest and the post-test.
Mean$_{\text{correct}}$: 3.684 - Mean$_{\text{incorrect}}$: 3.338; insistent request: $t(4)= 1.865$, $p = .136$, Mean$_{\text{correct}}$: 3.212 - Mean$_{\text{incorrect}}$: 6.376). Infants’ “correct” looking times were 54.42% for statement, 56.1% for yes-no questions and 37.29% for insistent requests.

**Figure 10.** Mean looking times (in seconds) for correct (matched) and incorrect (mismatched) stimuli (statement and vocative). Error bars represent standard error

**Figure 11.** Mean looking times (in seconds) for correct (matched) and incorrect (mismatched) stimuli (yes-no question)
A special analysis of trochees and iambs was also done on the insistent request condition, with “correct” looking times examined. A significant difference between looking times during “correct” and “incorrect” stimuli was found for neither trochees ($t(4) = -1.724, p = .160$) nor iambs ($t(4) = -0.591, p = .586$). Infants’ “correct” looking times were 22.24% for trochees and 45.05% for iambs (Mean-trochee$_{correct}$: 2.160 – Mean-trochee$_{incorrect}$: 7.552; Mean-iamb$_{correct}$: 4.264 – Mean-iamb$_{incorrect}$: 5.200).

**Figure 12.** Mean looking times (in seconds) for correct (matched) and incorrect (mismatched) stimuli (insistence request)

**Figure 13.** Mean looking times (in seconds) for correct (matched) and incorrect (mismatched) stimuli (insistent request-trochee)
Trials in which children looked less than 2 seconds or more than 19 seconds were excluded from analyses.

3.7.- Discussion

This pilot study (Experiment 2) taught two similar sounding words for two novel concepts with different intonations to 14-19-month-olds. Prior research has found that infants at this age have not yet perceptually mastered the intonational grammar of their native language, and our results tend to tend to confirm this. Though we cannot yet draw general conclusions due to the small sample analyzed in this preliminary study (we would need 16 infants to obtain reliable statistics), we would like to underscore two basic results:

a) The visual choice experimental design is useful for this kind of task.

b) There seem to be tendencies in the mean looking times which point to adult-like behaviour in infants as young as 14-19 months old.
The visual choice design has already been successfully used in children as young as 14 months old, and the results of our experiment point in the same direction. Our preliminary results show that, following a habituation phase, infants at this age can learn new similar-sounding words and learn to distinguish them. Curtin (2009) demonstrated that English 12-month-olds could learn new words differing only in their stress pattern. We have shown that Catalan and Spanish children also seem to be capable of doing this.

We can divide our preliminary results in three groups according to the target intonation patterns. First, we used in the habituation phase a statement intonation (the one usually used in linguistic experiments involving infants) mixed also with a vocative intonation. In the test phase we used the same intonation as in the habituation phase, and these are the results.

Although the difference in looking time is not statistically significant, we see that infants’ gaze is longer when the object seen matches the name heard than when there is a mismatch in stimuli (there was a difference of 1.542 seconds). Thus, we can hypothesize that Spanish and Catalan infants behave like English infants in being able to associate two stress conditions, trochee and iamb, to two newly learned words.

However, the main aim in this experiment was to evaluate whether infants would be capable of associating a variety of intonation contours with the newly learned word. We therefore modified the habituation stimuli by giving a second intonation pattern, that of a yes-no question. In this case, though infants’ looking times are not significant for matched relative to mismatched stimuli, the tendency, again, is for them to look longer at the correct word.

The third intonation was selected as a control intonation because it was shown in Experiment 1 to be ambiguous for adults. Our hypothesis predicted confusion in babies in these last test trials, since this is precisely what happened in the experiment with adults. Again, even though there is no significant difference in looking times, crucially
looking times in this case were higher when infants were looking at the visual image that did not match the word they heard.

We insist that our conclusions must be treated with caution (since, as we note above, such a small sample does not allow us to extract reliable results), but the fact that looking times for matched stimuli in the statement/vocative and yes-no question are 54.425% and 56.1%, respectively, yet only 37.29 in the case of the insistent request condition might not be mere chance. In fact, infants’ perceptual mechanisms might work the same as adults’ (despite the developmental factor, of course), so a misinterpretation due to duration or complex contour movement might equally well be at work in infants as it is in adults.

In order to better understand the expected behaviour elicited by insistent request contours, we divided the responses as a function of the stress condition, that is, we separated the data into trochees and iambs. Our results showed that infants had more difficulty (though the difference was small) with trochees, which were confused with iambs. This is exactly the same behaviour that was found in the adult experiment (Experiment 1). We hypothesized that the complex pitch boundary movement of the target L+H* LHL% contour on a trochaic word like Dlbu is inducing perceptual salience of the second syllable, even though the stress is on the first one. That is why some adults’ tended to confuse it with an iambic word, and, we hypothesize, infants at 1;6 might display the same behaviour.

4.- CONCLUSION

The goal of this study was to explore infants’ early intonational development. Little work has been done so far in this field, and we wanted to determine infants’ early abilities regarding the recognition of differing intonation patterns.

Intonation is an important phonetic feature which might be developed early in life. Thus, two experiments were designed to test our hypothesis. We first ran an adults’
experiment to assess the difficulties posed by some intonations in combination with two stress conditions (iamb and trochee). The f0 movements of intonation (which is one of the most salient cues for the perception of stress) can make the unstressed syllable more salient, thus leading to a certain amount of confusion. This first experiment showed us that statement, vocative and yes-no question intonations were the most easily discriminated, whereas some others, namely gentle request and insistent request intonations, were the most difficult ones to distinguish. In order to test infants’ abilities, we chose the three easier intonations and one of the complex ones (the insistent request) for the child experimental task.

The results of the second experiment point to in the same direction of Experiment 1. Though the sample used was not so large as to offer us a reliable conclusion, we can tentatively hypothesize that 14-19 month olds already have a (generally) stable intonational grammar. They are able to associate a particular contour with each stress condition, thus having learned the correlation between stress and intonation. Moreover, the infants in our study did not have difficulty with the most easily discriminable intonations, while the insistent request was very often misinterpreted, especially in the trochee condition, presumably due to the complex boundary tone movement.

Therefore, we can conclude that infants have an almost adult-like behaviour at 14-19 months of age. These results completely coincide with those of production studies, thus creating an important link between perception and production. However, more research is needed in this direction, so that our results can be tested against other languages.
5.- ACKNOWLEDGEMENTS

I would like to extend my heartfelt thanks to everybody that contributed in one way or another to this work. First and foremost, I am especially grateful to Professor Pilar Prieto, who has guided my work throughout this year and has always been available for my questions, even in summer. I would also like to thank Núria Sebastián for her valuable comments and willingness to let me use the infant laboratory, and Ferran Pons (and David), who patiently spent long hours with me during the entire process: setting up the experiment would not have been possible without his help. I am also indebted to Luca Bonatti, who rapidly solved my computer problems.

This work, however, is a reality thanks to the babies that were patient enough to watch our videos: Martina, Iu, Iker, Laura, Armand, Àlex, Mercè, Emma and Arnau. I am especially grateful, too, to their parents: only their interest in language acquisition made this study possible.

I am also greatly indebted to all those who participated in the adult experiment; to Joan Borràs (always there), for his technical and statistical support; to Xavi Mayoral, for his technical support; to Andrea Gargallo, for patiently spending her time (even when she did not have it) designing the pictures and videos; to Gisela Pi and Sara Sáez, because they made my work much easier and I felt at home with them; to Núria Esteve, who began this process with me and has always been available; to Ana Ruiz, for giving me the patience I do not have; and last, but not least, to Io Salmons, because our talks have always been so encouraging: working with her has truly eased my way down this long road.
6.- REFERENCES


