ACQUISITION OF PROSODY
AT THE BABBLING STAGE IN CATALAN

PROSODIC, GESTURAL, AND PRAGMATIC CONSIDERATIONS

Núria Esteve Gibert

Universitat Pompeu Fabra

Ph.D. Project
Supervisor: Dr. Pilar Prieto
0. ABSTRACT

Children develop the ability to communicate their emotions and intentions before they produce their first words (Bates, Camaioni, & Volterra, 1975; Piaget, 1936, 1946; Tomasello, Carpenter, & Liszkowski, 2007). Which are the tools that children use to this end? Previous work in the field suggests that babbling children produce vocalizations with distinct prosodic patterns depending on communicative context (D’Odorico & Franco, 1991; Marcos, 1987; Papaeliou, Mindakis, & Cavouras, 2002; Papaeliou & Trevarthen, 2006). Other studies have found that gestures are a crucial tool for children to communicate their emotions and intentions with parents or caregivers (Bates et al., 1975; Brinck, 2004; Camaioni, Perucchini, Bellagamba, & Colonnese, 2009; Cochet & Vauclair, 2010; Goldin-Meadow, 2007; Liszkowski, 2005; Tomasello et al., 2007).

The goal of this dissertation is to investigate the early acquisition of prosodic and gestural patterns by Catalan-babbling infants. Thus, three main issues will be addressed: (1) the way in which gestural and prosodic cues signal the emergence of intentional communication at the babbling stage, (2) the gesture-speech relation at the babbling stage, focusing on how children temporally align gesture and speech and comparing results with adult data, and (3) if prosodic cues are the tool that children use to understand the different meanings behind a pointing gesture. As far as we know, no previous research has studied the acquisition of prosody and gestures in Catalan-babbling infants.

In order to do this, the following data will be analyzed: first, a longitudinal corpus of four Catalan-monolingual children will be used to study whether children use gesture and prosody to communicate and also to investigate the temporal alignment between both modalities. Next, the results of an experimental task with adults will allow us to compare children’s results with adults’. Lastly, a perception task with an eye-tracking system will be used to investigate whether the parents’ use of specific request prosody while pointing to an object increases children’s comprehension of the speech act in comparison with the pointing gesture accompanied by neutral prosodic contours. This work, then, aims to offer a general perspective on the use of prosody and gestures from a perception and a production point of view. In sum, we believe that this dissertation will contribute to the study of language acquisition, giving evidence that there is a stage before the emergence of the first words when children have already developed very important linguistic skills.
RESUM

És sabut que els infants desenvolupen la capacitat de comunicar emocions i intencions abans de poder produir les primeres paraules (Bates, Camaioni, & Volterra, 1975; Piaget, 1936, 1946; Tomasello, Carpenter, & Liszkowski, 2007). Quines eines utilitzen per a fer-ho? Alguns estudis previs suggereixen que durant el balboteig, els nens produeixen vocalitzacions amb patrons prosòdics diferents segons la situació comunicativa en què es trobin (D’Odorico & Franco, 1991; Marcos, 1987; Papaeliou, Mindakis, & Cavouras, 2002; Papaeliou & Trevarthen, 2006). Altres autors han demostrat que els infants utilitzen els gestos com a eina per a comunicar les seves emocions i intencions amb els que els envolten (Bates et al., 1975; Brinck, 2004; Camaioni, Perucchini, Bellagamba, & Colonnesi, 2009; Cochet & Vauclair, 2010; Goldin-Meadow, 2007; Liszkowski, 2005; Tomasello et al., 2007).

L’objectiu de la tesi és investigar com els infants catalans adquireixen els patrons prosòdics i gestuals durant l’etapa del balboteig. Per això, ens centram en tres qüestions principals: (1) la manera com els gestos i la prosòdia demostren l’emergència de la comunicació intencional en els infants, (2) la relació entre gest i parla durant el balboteig, sobretot des del punt de vista de l’alineació temporal entre ambdues modalitats, tot comparant els resultats amb dades de parlants adults de català, i (3) la utilització de la prosòdia per part dels infants a l’hora d’entendre els diversos sentits dels gestos díctics.

Per tal d’acomplir aquest objectiu, s’analitzaran diversos tipus de dades. En primer lloc, l’anàlisi d’un corpus longitudinal de quatre infants catalans monolingües ens permetrà veure com els infants utilitzen gestos i prosòdia per comunicar-se i veure com s’adquireix l’alineació temporal de gest i parla durant el balboteig. En segon lloc, a partir de l’anàlisi dels resultats d’una tasca experimental amb adults, podrem comparar les dades amb els infants. Finalment, durem a terme una tasca de percepció amb eye-tracking per a veure si quan els pares assenyalen un objecte mentre produeixen un contorn entonatius de petició, els infants entenen més clarament la intenció de l’adult que no pas si el contorn entonatius és neutre. Pel que ens consta, mai abans s’havia estudiat l’adquisició de la prosòdia i els gestos en infants catalans. Creiem, doncs, que aquest treball servirà per mostrar com s’adquireix la prosòdia i els gestos en català, tant pel que fa a la producció com pel que fa a la percepció. En definitiva, pensem que podem provar que abans de produir les primeres paraules, els infants ja han sigut capaços de desenvolupar habilitats lingüístiques molt importants.
1. INTRODUCTION

1.1. Object of analysis

This research focuses on the study of the early acquisition of prosody by Catalan-speaking infants during the babbling stage. The study wants to analyze four main topics: (1) how intentional communication at the babbling stage is signaled by gestural and prosodic cues, (2) the development of gesture and speech before children are able to speak, focusing on their alignment, (3) the differences between adults and children in the temporal alignment between gesture and speech, and (4) how children who still do not speak are able to perceive the distinct meanings of gestures and whether prosody plays a role in this process of understanding others’ intentions.

1.2. Prior work

Studies on the acquisition of prosody have investigated the first speech patterns from different points of view. Some studies have focused on early perception of prosody, finding that attunement of stress perception abilities in the native language starts at 4-5 months of age, and that infants are aware of the stress properties of their mother tongues at around 9 months of age (Jusczyk, Cutler, & Redanz, 1993; Pons & Bosch, 2007; Weber, Hahne, Friedrich, & Friederici, 2004). Less is known about the early infant’s ability to discriminate between intonational contours. Sato, Sogabe, & Mazuka (2006) discovered that infants at 0;4, 0;6, and 0;8 could distinguish pitch patterns of target words, and Nazzi, Floccia, & Bertoncini (1998) found that newborns are able to extract pitch contour information at the word level.

1.2.1. Continuity between babbling and early words

As for the early production of intonational contours, many studies support the hypothesis of some kind of continuity between babbling and early words by finding language-specific influences in infants’ vocalizations (Boysson-Bardies & Vihman, 1991; DePaolis, Vihman, & Kunnari, 2008; Hallé, Boysson-Bardies, & Vihman, 1991; Levitt & Utman, 1992; López-Ornat & Karousou, 2005; Whalen, Levitt, & Wang, 1991). Cross-linguistic research has found, for instance, that French newborns tend to
produce rising melodies in early cries whereas Germans tend to produce falling contours (Mampe, Friederici, Christophe, & Wermke, 2009). Likewise, Levitt & Utman (1992) reported language-specific influences in one French-speaking child and one English speaking child at 0;5, 0;8, 0;11, and 1;2, since the French child produced non-final syllables closer in duration to one another than did the American infant. DePaolis et al. (2008) found language-specific differences when comparing intonation of 10 English-, 10 French-, 5 Finnish- and 5 Welsh-speaking infants at the 4-word stage.

Whalen et al. (1991) studied 5 French-speaking children and 5 English-speaking children from 0;7 to 0;11, finding that intonation patterns were different depending on the language environment: French children used more rising intonation and less falling intonation than English children. However, the authors also suggested that some of these differences in contour direction could also be due to communicative purposes. Similarly, Snow (2006) studied contour direction and pitch range of 60 English-learning infants between 0;6 and 1;11 and concluded that the acquisition of intonation follows these developmental stages: at 0;6-0;8 pitch patterns reflect pre-intentional mechanisms linking physiological states and emotional experience; at 0;9-0;11 there is a regression phase suppressing intonational expressiveness; and finally at 1;6-1;8 pitch patterns are actively controlled.

In general, these prosodic studies have not investigated prosody in relation to its pragmatic intentions. In the following section some results are presented about studies that have tried to link these two linguistic aspects, thus offering a wider perspective on the acquisition of prosody.

### 1.2.2. Development of intonation in relation to the development of intentional communication

Some studies analyzing intonational contours of babbling infants have suggested that the development of intonation has to be studied in relation to the development of intentional communication.

To our knowledge, the first author to study the emergence of communicative acts in that sense was Halliday (1975). He analyzed his son’s early pitch contours from 0;9 to 2;6 and discovered that different vocal expressions were able to convey distinct functions. Halliday found that his child produced mid falling tones when interacting with
other people but low falling tones with narrower range when he was interested in the modification of an object. López-Ornat, Gallego, Gallo, Karousou, Mariscal, & Nieva, (2003) studied vocalizations from 0;8 to 2;6 in 95 Spanish-speaking infants. Their results revealed that children as young as 0;8 were able to produce vocalizations with pragmatic functions such as proto-conversations (defined as the child producing turn-taking dialogues, with or without recognizable words) or proto-declaratives (defined as the child trying to attract someone’s attention about something (s)he likes and accompanying it with gestures).

D’Odorico & Franco (1991) studied 5 Italian-learning children from 0;4 to 0;11 months of age, analyzing acoustically their vocalizations in terms of mean f0 values, maximum and minimum pitch, average number of pitch change and units of vocalizations in a prosodic unit, and mean duration. As for context types, vocalizations were classified as vocalizations during infant manipulation of a toy (VIM), vocalizations during shared experience (VSE, i.e. manipulating a toy but looking at the adult), vocalizations during adult manipulation of a toy (VAM), and vocalizations during exchanges with the adult (VEA, i.e. neither of them is manipulating the toy but they are both looking at each other). Results offered support for a ‘selective production hypothesis’ whereby different types of vocalizations were produced in different communication contexts until children were 0;9. Thus, children at 0;4-0;6 used different contour directions when producing a VIM and a VSE; at 0;6-0;8 children assimilated categories VSE and VAM; and at 0;8-1;0 VIM vocalizations could not be distinguished from the other vocalizations. The authors hypothesized that a child’s ability to acoustically distinguish between categories tends to disappear as age increases. Therefore, children show a selective production hypothesis until 0;9 but not thereafter, i.e. different patterns of non-segmental features characterize sounds produced in different contexts. Because their results revealed many individual differences among their infant subjects, the authors concluded that they had failed to capture communicative differences across contexts.

In order to study how children use prosody to express distinct meaning, Marcos (1987) analyzed 10 English-learning children between ages 1;2 and 1;10, i.e. already at the one-word period. Results suggest that pitch range is higher for repeated requests than for initial requests for objects, and higher for initial requests for objects and for cooperation than for labeling. The study also analyzed vocalizations when the child was giving and showing an object and found that their pitch range is in an intermediate rank between requests and labeling.
Papaeliou, Mindakis, & Cavouras (2002) analyzed six English-learning children from 0;7 to 0;11 and found that prosodic patterns were different when vocalizations conveyed communicative functions from when they expressed emotions: vocalizations carrying communicative functions were shorter, with lower f(0) values, and had greater intensity than vocalizations expressing emotions. Similarly, Papaeliou & Trevarthen (2006) observed that children displayed different prosodic patterns when vocalizations were classified as communicative, or when classified as investigative: compared to investigative vocalizations, communicative vocalizations had a higher mean and maximum f(0), higher standard deviation of f(0), and shorter duration.

In sum, research on how children use prosody in order to express pragmatic meanings suggest that at a very early stage in language development, children vocalize with different prosodic patterns according to the context involved in the communicative situation. However, very few studies investigate this hypothesis in children that are still not producing words, i.e. at the babbling stage. And those studies focusing on this early period have mainly found that babbling children use distinct prosodic patterns to distinguish communicative from non-communicative (investigative) vocalizations. (Papaeliou et al., 2002; Papaeliou & Trevarthen, 2006). Nevertheless, more research is needed to investigate if babbling children use prosody as a tool not only to indicate the communicative status of the vocalizations but also to express specific pragmatic meaning, since it is during the babbling period that the emergence of intentionality occurs (see section 3.2).

1.2.3. Communicative use of gestures

At the babbling stage, however, children communicate not only through prosody. At this stage their use of gestures is crucial to communicate their emotions and intentions with parents or caregivers. Research on the communicative role of gestures is based on the assumption that human communication is inferential (Grice, 1957; Sperber & Wilson, 1986), which means that (1) interlocutors form intentions toward others’ intentions and understand epistemic states to transmit and infer referential content, and that (2) human communication is cooperative, i.e. the sender wants the recipient to understand the message, and the recipient wants to understand the message.

Another state of affairs is the case of children's use of gestures. Right after birth the gestures produced by children are not communicative. However, at the age when
children have completely developed the theory of mind, i.e., around four years of age, they use gestures as communicatively as adults (McNeill, 1992). However, there is no consensus about the communicative status of gestures during the period ranging from the development of intentionality (around 0;9) and 4;0. Studies investigating the communicative use of gestures in children have found results in two main directions. From the one hand, infants’ gestures have been interpreted as a social tool that builds on the emerging intentionality of children (Bates, Camaioni, & Volterra, 1975) and that would only serve one’s own benefit in an individualistic problem-solving way. According to this view, then, it would not be clear if children gesture to influence the others’ intentional states in order to benefit the other and with cooperative motive to communicate. From the other hand, infants’ communication is seen as a cooperative tool from an ontogenetic point of view (Bruner, 1983; Werner & Kaplan, 1963) and a motive for social contact in an ultra-social environment. This perspective, however, has to face the issue of whether infants’ behavior is communicative or whether adults interpret this behavior as communicative.

Liszkowski (2008) suggests a social-cognitively and motivationally rich referential communication in children before they are able to speak. His research focuses on infant deictic gestures, i.e. giving, showing, reaching, and pointing. At around 0;9, children give and show objects to the others and they seem to cooperate and bring objects to the attention of others. At this age, however, it may be that children are only interacting with others in a non-referential way, simply with the purpose of establishing social contact. Yet, at 1;0 children refer to present or absent entities by pointing, indicating that they understand that others will attend to their gesture, that they can attract their attention, and understand the shared background. Many other studies have found evidence for social-cognitive and cooperative motives in infant pointing (Brinck, 2004; Camaioni, Perucchini, Bellagamba, & Connesni, 2009; Cochet & Vauclair, 2010; Goldin-Meadow, 2007; Liszkowski, 2005; Tomasello, Carpenter, & Liszkowski, 2007). Brinck (2004) distinguishes between declarative pointing –performed in order to make the addressee do something for the subject– and imperative pointing –an attention-seeking device to achieve joint reference with the addressee–. The author suggests that only imperative pointing has a social motivation for cooperation, while declarative pointing is mainly intersubjective. Contradictory results were found by Camaioni et al. (2009), whose results suggest that declarative pointing is linked to the understanding of others’ intentions, and that it emerges later than imperative pointing. Tomasello et al. (2007) suggest that infants point in a social-cognitive and cooperative way, since they
aim at directing others’ attention to entities within a joint attention frame, even with absent entities. They distinguish between three types of pointing gestures:

- **Informative (assertive) pointing**: the communicator wants the recipient to know something that he thinks she will find useful or interesting.

- **Requestive (directive) pointing**: the communicator wants the recipient to do something that will help him, the communicator, in some way (including by providing needed information, as in questions).

- **Expressive pointing**: the communicator wants the recipient to feel some attitude or emotion that he is already feeling.

Cochet & Vauclair (2010) investigated differences between imperative pointing, declarative expressive pointing, and declarative informative pointing. They found that different hand shapes were used depending on the pointing gesture: imperative gestures were associated with whole-hand pointing and declarative gestures were more frequently characterized by an extended index finger. In terms of the speech accompanying the gesture, declarative gestures were more frequently accompanied by vocalizations than imperative gestures.

Pointing gestures are broadly considered the first gestures used by children with a communicative purpose. Adults, nonetheless, use other gestures to communicate apart from the deictic ones. However, the debate on whether gestures and language form a single system is also very present in the literature on adult gestures. One perspective suggests that gestures assist communication and do not provide any meaning apart from that provided by speech (Butterworth & Beattie, 1978; Butterworth & Hadar, 1989; Hadar & Butterworth, 1997; Krauss, Chen, & Gottesman, 2000). Another perspective suggests that gestures are both communicative and production-aiding. Goldin-Meadow (2005) states that gesture without speech takes a language-like form (with sentence-level structure, word-level paradigm, and grammatical categories), whereas co-speech gesture loses the language-like form and helps in conveying thoughts that do not easily fit into the categorical system of language. Kita (2000) proposes the “Information Packaging Hypothesis”, stating that some gestures help speakers organize information in packages suitable for expression in a single utterance, since gestures also access to a spatio-motoric way of thinking.

Finally, many studies have found evidence in favor of the hypothesis that gesture is communicative and that it provides meaning apart from that provided by speech (Clark,
1996; Duncan, 1996; Kendon, 2004; Loehr, 2004; McClave, 1991; McNeill, 1985, 1992). For instance, Bavelas, Kenwood, Johnson, & Phillips (2002) found that social context in the form of visibility between speaker and addressee influences gesture rate; Bavelas, Gerwing, Sutton, & Prevost (2008) and Gullberg (2006) found that social context influence aspects of gesture form; Furuyama (2000) and Özyürek (2002) discovered that the addressee location can influence speakers’ use of gesture space to represent semantic information, and Holler & Wilkin (2009) found that speakers accompany a higher proportion of definite references with iconic gestures than they do with indefinite references, that ‘action gestures’ (gestures foregrounding information about actions) accompany mainly definite references whereas ‘entity gestures’ (gestures foregrounding information about entities) mainly indefinite ones. McNeill’s (2005) work states the existence of a growth point, i.e. a minimal unit of an imaginary-language dialectic, a package that has both linguistic categorical and imagistic components. Five main reasons are given in McNeill (1992) in order to justify that gesture and speech form a single system: (1) gestures occur with speech in 90% of the cases; (2) gesture and speech are semantically and pragmatically coexpressive; (3) gesture and speech are synchronous; (4) gesture and speech develop together in children, and (5) gesture and speech break together in aphasia.

In conclusion, many studies have found that children use gestures to communicate their intentions. As far as we know, however, there are no studies dealing with the children’s use of gestures together with speech to convey their intended meanings. Some studies have claimed that gesture and speech form a single system. Do gesture and speech form a single system from the beginning? Do these two modalities develop together in children? Is there a specific point in time in children’s development where gesture and speech start behaving like a single system? More research is needed in order to solve these questions.

1.2.4. Synchronization between gesture and speech

Most of the research on the integration of gesture and speech, then, suggests a certain degree of communicative status of gestures. The synchronization between gesture and speech is used as one of the main arguments in favor of this hypothesis. The synchronization between the two modalities can occur at three different levels (Kendon, 1980, McNeill, 1992):
• **Semantic synchronization**: if gesture and speech co-occur they must refer to the same semantic unit.

• **Pragmatic synchronization**: if gesture and speech co-occur they perform the same pragmatic functions.

• **Phonological synchronization**: the stroke\(^1\) of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech.

According to the phonological synchronization rule, there is a close relationship between gesture and prosody, since the stroke of the gesture is generally synchronous with the pitch accent of the accompanying speech. Some studies have found evidence for the synchronization of gesture and speech at the phonological level in adult speakers. Ferré’s (2010) results, for instance, confirm that the stroke of the gesture starts before and ends after the semantically related word. Giorgolo & Verstraten (2008) state that synchronization between audio and visual material influences the addressee to consider the message as a good multimodal utterance or not, and this decision is processed according to prosodic and semantic criteria. Loehr (2007) found a rhythmic relation of hands, head, and voice such that each articulator produces peaks synchronically with the other articulators. Finally, Rochet-Capellan, Laboissière, Galván, & Schwartz (2008), discovered that arm movements and jaw movements are related temporally, since the pitch accent occurs synchronically with the stroke of the gesture, both in trochees and in iambs.

Very little is known about the development of the synchronization of gesture and speech in children. It seems that at early stages of language development, children show a preference for communication in the gestural modality over the verbal modality (Acredolo & Goodwyn, 1985, 1988; Iverson, Capirci, & Caselli, 1994). As for the semantic synchronization, it has been found that most of the gesture-speech combinations contain gestures that are redundant with the information conveyed in speech (Greenfield & Smith, 1976). However, these are not the total occurrences of gesture-speech combinations: in some cases, children’s gestures can convey information that is not present in speech, such as pointing at a ball while saying ‘give’ (Butcher & Goldin-Meadow, 2000; Capirci, Iverson, Pizzuto, & Volterra, 1996; Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005).

---

\(^1\) *Stroke: the peak of effort in the gesture. For more information, see section 5.2.2.*
As far as we know, Butcher & Goldin-Meadow (2000) is the only study that has dealt with the developmental point at which children behave like adults in the production of co-speech gestures, both in terms of the total proportion of gestures produced and in terms of the phonological synchrony between gesture phases and prosodic patterns. The authors analyzed 6 children from the one-word stage (14.5 months, on average) until the two-word stage (23.5 months, on average) and found that the production of gesture with respect to the total number of communicative acts is around 20% and remains stable across the stage. Their results also suggest that it is at the end of the one-word period that children produce gestures mainly in combination with speech, such as adults do. In terms of the temporal alignment, the authors considered that gesture and speech were adequately aligned when the vocalization occurred on the stroke of the gesture\(^2\) or at the peak of the gesture\(^3\). They found that the synchronization of gesture and speech does not occur until the end of the one-word period. In terms of the semantic synchronization, they found that the combination of communicative gestures and meaningful words (as opposed to meaningless vocalizations, i.e. vocalizations that do not refer to any target word) increases a lot during the one-word period. In conclusion, the authors found that it is when gesture is combined with ‘meaningful speech’ that the two modalities are phonologically synchronized.

Previous work, therefore, suggests that babbling children do not synchronize prosody and gesture, since it is not until children produce two-word combinations that they start doing so. However, no research has been actually done to investigate how the two modalities are aligned at the babbling stage. It may be that phonological synchronization is still not adult-like, but we no study has shown how it is. More research is needed to describe the temporal alignment between prosody and gesture at a stage when children use both modalities to convey their intended meanings.

1.2.5. Intonation as a tool for children to determine intentions in co-speech gestures

---

\(^2\) Stroke: the peak of effort in the gesture. For more information, see section 5.2.2.

\(^3\) Peak of the gesture: the farthest extension before the hand began to retract. It coincides with the final point in the stroke. Other authors such as Loehr (2004) have called this point ‘apex’.
In his development of the Cooperative Principle, Grice (1975) stated that humans communicate mainly with the motivation of being cooperative. In fact, previous work by Austin (1962) and Searle (1969) in the ‘theory of speech acts’ had established five main types of speech acts present in any communication act: assertives, directives, commissives, expressives, and declaratives (see section 3.2). From these five speech acts, Tomassello et al. (2007) consider that three of them are the most basic and defined in terms of helping and sharing:

- **Assertive** (or informative): the communicator wants the recipient to know something that he thinks will be useful or interesting.
- **Directive** (or requestive): the communicator wants the recipient to do something that will help the communicator in some way.
- **Expressive**: the communicator wants the recipient to feel some attitude or emotion that the communicator is already feeling.

The authors state that when children are in their first stages of language development, their motivations when communicating are mainly of these three kinds: assertive, directive, and expressive.

There is some evidence that children can understand these motivations at an early age in language development. Some research has found that around 3;0, children can infer the intention hidden in an adult communicative act (Povinelli, Reaux, Bierschwale, Allain, & Simon, 1997; Tomasello, Call, & Gluckman, 1997). And even earlier in language development, Behne, Carpenter, & Tomasello (2005) found that at 1;2, children can already infer the location of a hidden toy by means of the communicative cues they observe in adults. However, they also reported that children could infer the hidden location only when the adult produce a clear cue such as ostensive gazing or pointing to the location, not when the adult gaze to the object in an absent-mindedly way.

In a study of children from 0;9 to 1;5, Camaioni et al. (2009) found that comprehension precedes production of communicative pointing gestures in typical development. Their results suggest that children produce earlier imperative pointing than declarative pointing, and that this difference is due to the fact the both types of pointing entail different socio-cognitive abilities. They argue that imperative pointing is just an understanding of others as causal agents, whereas declarative pointing reflects an understanding of others as intentional agents. They also found that children who first
inferred the act of another person were also the first to produce declarative pointing at 1;0 and at 1;5. However, no relation was found between the production and comprehension of imperative pointing and intention understanding.

Aureli, Perucchini, & Genco (2009) analyzed children from 1;4 to 1;8 and found that they are able to recognize two different social intentions in a pointing gesture, depending on the common ground involved. Thus, children interpreted a pointing gesture as informative when the adult had pointed at the hidden toy’s location; however, children identified a pointing gesture as declarative when the adult pointed at a referent in a different context, since the child only reacted by commenting or smiling at the adult.

Very few studies address the relationship between gesture and prosody. In a recent study, Cochet & Vauclair (2010) tried to elicit imperative, declarative expressive, and declarative informative pointing in infants aged 1;3-2;6. They found that declarative pointing was accompanied by vocalizations more often than imperative pointing and that declarative pointing lasted longer than the imperative one. They state that it would be interesting to investigate the nature of the vocalizations accompanying declarative and imperative pointing, suggesting that imperative pointing might occur more often with vocalizations, whereas declarative pointing might occur with words (or pseudowords). The authors also suggest that imperative pointing could be originated from the non-communicative reaching actions, since children might learn that their reaching actions produce an effect on adults. However, they hypothesize that declarative pointing might be originated from imitation processes.

Thus, it is seems reasonable to assume that children are able to produce pointing gestures with distinct social motives at an early age in language development. And that even before in their developmental process, they have the ability of understanding these motives in others. Thus, some studies suggest that babbling children understand the intention behind an adult’s pointing gesture by means of the common ground involved and other cues such as ostensive gazing. However, more research is needed to investigate the role of prosody in this early ability in cognitive development. Since most of the adults’ gestures are accompanied by speech, it might be that children’s rely on this speech to understand motivations in gestures.

In conclusion, previous work on the use of prosody to communicate intentionally has suggested that at the babbling stage, children use prosodic contours to distinguish
vocalizations uttered with a communicative purpose, with an emotive purpose or without any of them. However, as far as we know, no previous study has dealt with this issue with a deeper analysis: do babbling children use prosody not only to signal the communicative status of their vocalizations but also to express specific pragmatic meanings within communicative vocalizations? Previous literature has also shown a tight relation between gesture and prosody at early stages in language development, suggesting that children use both modalities to communicate, sometimes together and sometimes separately. Yet, more research is needed to know when children start using prosody and gestures as a single system, such as adults seem to do. And the broad picture of the use of prosody—a accompanied or not by gesture—to communicate would be incomplete if we do not deal with it from a perceptual point of view. Thus, we would be sure about the abilities that children have already acquired at this early stage in language development.

Given that some questions still remain unsolved in the previous literature, this dissertation aims at addressing them, both from a production and a perceptual point of view.

2. GOALS OF THE DISSERTATION

This work has the goal to investigate the early acquisition of prosody by Catalan-speaking infants during the babbling stage. Three main issues will be addressed: (1) the way gestural and prosodic cues signal the emergence of intentional communication at the babbling stage, (2) the integration of gesture and speech at the stage when children are still not able to speak, focusing on the alignment of the two modalities, and comparing results with adult data, and (3) if prosodic cues are a tool that children use to understand the different meanings behind a pointing gesture.

To our knowledge, no previous research has studied the acquisition of prosody in Catalan-babbling infants, even though prosody is accepted to be one of the first linguistic abilities to be acquired in the language development process (Prieto et al., in press). Crucially, before children are able to speak, they have developed intentional communication (Bates et al., 1975; Piaget, 1936, 1946; Tomasello et al., 2007). Then, do babbling children express somehow their recently acquired intentions? Our hypothesis is that before children are able to produce words, prosodic cues and gestures are the tools they use to communicate emotions, intentions, and needs. This
study also aims at investigating if the relation between gesture and prosody at the babbling stage occurs in an adult-like manner in terms of the temporal alignment. Finally, this study also wants to investigate the role of prosody in determining the specific intentionality behind a pointing gesture. Thus, this work will try to offer a general perspective on the use of prosody and gestures from a perception and a production point of view at an early stage in language development.

3. RELEVANT THEORETICAL FRAMEWORK

3.1. The study of prosodic development

The Autosegmental Metrical model of intonation (which started with Pierrehumbert, 1980) is one of the most widely used frameworks to analyze prosodic patterns. This framework represents intonation in terms of level pitch targets, and annotates pitch movements associated with metrically strong syllables (pitch accents) and boundary tones (prosodic boundaries) by means of the Tone and Break Indices system, or ToBI. Some studies on the acquisition of prosody have successfully used the Autosegmental Metrical model to describe the intonation system in children (Frota & Vigário, 2008; Prieto, Estrella, Thorson, & Vanrell, in press).

Once children have acquired their first words, i.e. at the 25-word period, an analysis in terms of pitch accents and boundary tones in AM terms has been proposed. Frota & Vigário (2008) found that a European Portuguese child acquired the inventory of pitch accents and boundary tones in an adult-like way at 1;9, with the emergence of such contours as early as 1;5. For this child, intonational development occurred five months before the onset of the two-word stage, which was at 2;2. Prieto et al. (in press) investigated the development of prosodic patterns in four Catalan children and in two Spanish children and demonstrated that children at 1;1 and 1;3 are able to produce a set of adult-like intonation contours: at the one-word period, children produce statements and calling contours, and at the two-word period they produce complex contours such as requests, counter-expectations, (L* HL%), insistent requests (L+H* LHL%), and interrogative contours (L* HH%). Their results, thus, indicate that Catalan and Spanish children produced the basic phonologically distinct pitch contours from the 25-word period, and that at the two-word speech they show an important knowledge of the adult intonational grammar.
However, the AM model cannot be used when studying prosodic patterns at the babbling stage because children do not produce clearly defined strong syllables in their first vocalizations. Hence, research on prosody at the babbling stage has to study suprasegmental issues other than those involving only pitch accents and boundary tones. Typically, the development of prosodic patterns before children can produce pitch accents and boundary tones has been analyzed mainly through the following phonetic distinct features: pitch direction, pitch range, and duration (DePaolis et al., 2008; Snow & Balog, 2002).

Pitch direction (or contour direction) is the curve that tracks the perceived pitch over time. It indicates intonation and it can be produced in two distinct directions: rising intonation (the pitch increases over time) or falling intonation (the pitch decreases over time). Pitch or contour direction has been widely used in cross-linguistic studies in order to compare children’s productions with adult’s productions in languages with a typical contour direction. For instance, some studies comparing French and English have observed that children follow adult-like patterns in producing their vocalizations with a rising or falling contour, respectively (DePaolis et al., 2008; Vihman & DePaolis, 1998; Whalen et al., 1991).

Pitch range is defined as the existing interval between the valley and the peak in a rising pitch movement, or between the peak and the valley in a falling pitch movement (Prieto, 2003). Pitch range has been studied in early children’s productions to investigate whether children control it in an adult-like manner, i.e. if they use it to distinguish between semantic meanings of the utterances. Marcos (1987), for instance, analyzed 10 English-learning children between ages 1;2 and 1;10. The author found that pitch range is higher for repeated requests than for initial requests for objects, and that it is higher for initial requests for objects and for co-operation than for statements. The study also analyzed giving and showing vocalizations and found that their pitch range is in an intermediate rank between requests and statements. Snow (2006) suggests that the magnitude of children’s accent range develops markedly at about 18 months of age and proposes that the pattern of intonation development is U-shaped: at 0;9-0;1 there is a decline in accent range production that is equal in magnitude to the significant increase that occurs at 1;6-1;8. Esteve-Gibert & Prieto’s (submitted) also show that children use wider pitch range when producing communicative vocalizations than when not intending to communicate. The authors also found that when communicating, children use pitch range to distinguish between distinct pragmatic and
expressive intentions: pitch range was wider when requesting and expressing discontent, and narrower when responding to a stimulus or uttering a statement.

Many studies on the development of prosodic patterns have also analyzed duration in children’s vocalization. Levitt & Utman (1992) found that a French child produced non-final syllables closer in duration to one another than did the American infant. This result is consistent with adult French, since French has greater isosyllabicity in non-final syllables. DePaolis et al. (2008) found that duration showed the clearest difference across infants learning different languages such as French, English, Welsh, and Finish, since children exaggerate final syllable lengthening if the languages of their linguistic environments exhibit this phenomenon (as is the case of French). Hallé’s et al. (1991) results show that French children produced final lengthening on the last syllable of prosodic groups or words whereas Japanese children did not produce final lengthening, again in keeping with their ambient languages. Papaeliou & Trevarthen (2006) analyzed four English-learning infants at 0;10 and found that communicative vocalizations had shorter duration than investigative vocalizations. And Esteve-Gibert & Prieto’s (submitted) results confirm Papaeliou & Trevarthen’s hypothesis, since Catalan-babbling children produce longer vocalizations when communicating than when not communicating. The authors also found that when communicating, children use duration to distinguish between distinct pragmatic and expressive intentions: requesting and expressions of discontent lasted longer than responses to a stimulus or statements.

3.2. The study of intentional communication

The ‘speech act theory’ was first proposed by Austin (1962) and further developed by Searle (1969). This theory states that every sentence that is uttered in a language carries out these speech acts: locutions, illocutions, and perlocutions. Locutionary acts are acts required for the making of speech, i.e. saying something (a locution); illocutionary acts are conventional social acts recognized as such by both the speaker and the hearer and that take place when a sentence is uttered, i.e. the performance of an act in saying something; finally, perlocutionary acts are acts that have an effect, planned or unplanned, on the feelings, thoughts or actions of either the speaker or the listener. In Searle’s (1969) further development of the speech act theory, the author stated that all speech acts can be divided in two: a propositional content (or locution),
and a performative content (or illocution). The performative or illocutionary acts are mainly divided into one of these categories:

- **Assertives**: statements describing a state of affairs in the world.
- **Directives**: statements to make the other’s actions fit the propositional content.
- **Commissives**: statements to commit the speaker to a course of action.
- **Expressives**: statements expressing the ‘sincerity condition’ of the speech act.
- **Declaratives**: statements attempting to change the world by “representing it as having been changed”.

Besides, according to the Grice’s Cooperative Principle (1975), all communicative acts intend to fulfill two intentions: a communicator wants not only to convey a message but also that the addressee recognizes his intention to do so, and would not be satisfied with the outcome of his action unless this recognition is realized. Sperber & Wilson (1995) stated two main intentions: the communicator’s informative intention, aiming to achieve an effect in the addressee by modifying the cognitive environment, and the communicative intention, when the communicator aims to make his informative intention manifest for the addressee (or mutually manifest between the communicator and the addressee).

In general, research on language development has applied Searle’s distinction between propositional content and performative content, finding that children provide performative content in their vocalizations before the propositional content is adequately displayed. As Bates et al. (1975) point out, in children a locution requires the uttering of sounds and construction of propositions, thus implying the onset of verbal speech, an illocution requires the intentional use of a conventional sign to carry our any socially recognized function, and a perlocution simply requires that the signal produced by a person have some effect on the recipient. Merging Austin’s and Searle’s proposal, Bates et al. (1975) propose the following developmental stages:

- **A perlocutionary stage**: the child has a systematic effect on his listener without having an intentional control over that effect.
- **An illocutionary stage**: the child intentionally uses nonverbal signals to convey requests and to direct adult attention to objects and events.
- **A locutionary stage**: the child constructs propositions and utters speech sounds within the same performative sequences that he previously expressed nonverbally.
Research on development of intentional communication has focused on two main performative acts (Bates et al., 1975): imperatives and declaratives. Imperative acts use the adult as the means to a desired action or object, whereas declarative acts use an object (through pointing, showing, giving, etc.) as the means to obtain adult attention. This use of human agents to obtain or operate objects, and the use of objects to operate on human attention appears at around 0;8-0;10 months of age, and represents the children’s development of illocutionary acts. It starts with actions like showing and it progressively changes into giving and pointing actions. This period of development of the illocutionary acts coincides with the fourth and fifth stages in the sensorimotor stage of the Piaget’s theory of cognitive development (Piaget, 1936, 1946). At the fourth and fifth stages (going from 0;8 to 1;8), children begin to plan deliberately the steps to meet an object and discover new means to meet goals.

Hence, in the study of the development of intentional communication the concept of the theory of mind plays a central role. The theory of mind is the ability to attribute mental states, i.e. pretending, desires, intents, belief, knowledge, to oneself and others, and to understand that others have mental states that are different from one’s own. When studying the cognitive development of the theory of mind in children, one of the aims of the researchers is to establish the precise point of time in development where it is possible to identify in children the precursors of the theory of mind. Meltzoff, Gopnik, & Repacholi (1999) offer a review of the some findings in the development of the theory of mind. As Camaioni et al. (2009) state, it seems that children have developed the theory of mind at some age between 3 and 5 years old, but there is no strong evidence about what occurs before and during this stage. Bellagamba & Tomasello’s (1999) suggest that at 1;6 children can infer the goal of another person after seeing his/her unsuccessful attempts, but that at 1;0 children cannot imitate unsuccessful goal-directed actions. They claim that at 1;6 children understand that people’s actions are goal directed and intentional. Tomasello et al. (2007), Tomasello, Carpenter, Call, Behne, & Moll (2005), and Liszkowski (2005, 2006) propose a rich interpretation of the prelinguistic communications, suggesting that infant pointing for an adult gives evidence that the child tries to influence the adult’s intentional/mental state.

It is evident, then, that human interactions are produced on a common ground basis that enables the understanding of intentions in communication. This common background is the joint attentional frame, a triadic situation involving two people and an object or event outside the two. As shown by Behne et al. (2005), at 1;2 infants inferred a hidden toy’s location based on the common ground that both the infant and the adult
knew that the infant was seeking the toy. Instead, it has been reported that apes fail to perform the task of finding the hidden toy (Call & Tomasello, 2005). Most of the evidence found in favor of the early ability of children to establish joint attentional frames comes from the fact that before 1 year of age, children follow adult pointing gestures to targets and check back to make sure of her target (Carpenter, Nagell, & Tomasello, 1998). As explained in section 1.2.5, at 1;0 children point to establish social interaction and cooperation with three main motives (Tomasello et al., 2007):

- They want others to feel things (expressive pointing)
- They want others to know things (informative pointing)
- They want others to do things (requestive pointing)

Thus, at the end of their first year of life infants begin to use communicative gestures to actively direct adult attention to outside entities in triadic interactions. First and most important are deictic gestures such as pointing, showing, and offering, usually accompanied by the infant’s looking at the adult in alternation with looks to the object. The alternation of gaze between the object and the adult indicates that the child is aware of the effects his or her signals will have on the other person (Bates et al., 1975; Bretherton, 1991). As Tomasello & Camaioni (1997) documented, the key difference between human infants and chimpanzees is that infants use their deictic gestures — and especially pointing— not only for imperative purposes (to obtain a desired object or event) but also for declarative purposes (to share with another person interest or attention to some object or event).

In sum, understanding the common background or joint attention is a prerequisite for the development of the theory of mind and thus the adequate perception and production of communicative intentions. Even though at 0;3-0;6 infants point towards objects, their action is not communicative because they still do not understand intentions, attention, and shared attentions. However, at 0;9 children already understand goals and at 1;0 they understand perceptions. At these stages, too, infants begin to use communicative gestures to actively direct adult attention to outside entities in triadic interactions by means of deictic gestures such as pointing, showing, and offering. After children’s first birthday, i.e. 1;0-1;3 children can determine what others know, and it is also at this age that children are aware of the information they share with another person in a joint attentional frame.
3.3. The study of gesture development

Based on the assumption that gesture is communicative and that it forms a single system together with speech, and that in fact they are part of speech, McNeill (1992, 2005) defines gestures as ‘everyday occurrences —the spontaneous, unwitting, and regular accompaniments of speech that we see in our moving fingers, hands, and arms. The classification of the types of gestures proposed by McNeill (1992) is the most used in gesture studies. The author states that gestures can be classified into one of these categories (although many gestures involve more than one category):

- **Iconic gestures**, when there is a close formal relationship to the semantic content of speech.
- **Metaphoric gestures**, when the pictorial event presents an abstract idea.
- **Beats**, when the hand moves along with the rhythmic pulsation of speech.
- **Cohesive gestures**, used to tie together thematically related but temporally separated parts of the discourse.
- **Deictic gestures**, such as pointing, indicating objects and events in the concrete or abstract world.

Right after birth, however, human beings do not gesture in a communicative way. It is not until 0;8-0;10 that children gesture in order to influence the mental state of others, either because they want the others to do, know, or feel something. The first communicative gestures that typically developing children produce are deictic gestures such as pointing, giving, showing, or requesting (Bates et al., 1975; Iverson & Goldin-Meadow 2005; Özçalışkan & Goldin-Meadow, 2005; Sansavini, Guarini, & Stefanini, 2010; Tomasello, 2007). In these first gestures that infants produce, the referents can be identified only in the physical and social context in which communication takes place. At this age, infants produce another type of gestures: gestural routines such as clapping hands, kissing with the moving the hand from the mouth towards the recipient, and waving good-bye. At around 1;0, children start reproducing actions associated with specific objects, such as bringing a phone to the ear or brushing to the hair. They are called object-actions (Sansavini et al., 2010). Capirci, Contaldo, Caselli, & Volterra (2005) suggest that object-actions are later expressed in a symbolic way with iconic gestures. Thus, children would refer to the telephone by gesturing as having a phone by the ear and being in a conversation, showing that object-actions are then converted into iconic gestures. When focusing on the relation between gesture and speech at this stage, i.e. before the one-word stage, some research has suggested that children
produce more gesture-alone acts than gesture-speech combinations (Butcher & Goldin-Meadow 2000; Esteve-Gibert & Prieto, 2011). At the early one-word stage, however, this tendency is inverted and children already produce more gesture-speech combinations than gesture-alone acts. When looking at the gesture-speech combinations that children produce around the first year of life, it has been found that gesture and speech either convey the meaning of the same referent (for instance, producing a bye-bye gesture and saying 'bye') or the gesture identifies the referent and the word refers to its meaning (for instance, pointing to a dog and saying 'dog') (Capirci et al., 1996; Pizzuto & Capobianco, 2005).

From 1;4 to 1;8, children mostly produce pointing gestures. But they also produce iconic gestures, which at this age start appearing more and more with verbs and adjectives (Capone & McGregor, 2004). In terms of the relation between gesture and speech, Özçalişkan & Goldin-Meadow (2005) found that at 1;6 many children produce constructions combining argument + predicate in gesture-speech combinations (for instance, saying ‘mommy’ and pointing at a couch), whereas constructions like these do not occur in speech-only acts until children are 1;10. Similarly, at 1;10 many children produce utterances with two predicates (for instance, saying 'I like it' and producing an eat gesture) in gesture-speech combinations, whereas only few of them produce these kind of utterances only by means of speech.

Children increase significantly their usage of iconic gestures between 3;0 and 5;0, and their gesture-speech combinations are more adequately synchronized (Sansavini et al., 2010). At this stage, however, children still do not produce neither metaphoric gestures nor beat gestures. It is it not until children are 5 years of age that they develop completely the rest of the gestural system (McNeill, 1992).

The study of prosodic, intentional, and gestural development at early stages in language development has shown evidence that very young infants are able to communicate through gestures or prosody. In the present dissertation, nevertheless, we would like to demonstrate that these two modalities —namely, gesture and prosody— are very tightly related from the beginning, enabling children to communicate without words. This research would be a step further in the investigation in language acquisition, suggesting that babbling children use prosody and gesture as two faces of the same coin.
4. HYPOTHESES

Based on the research exposed in the previous sections, three main hypotheses will be tested in three experimental studies:

1. **Experimental study 1.** Gesture and prosody are the communicative cues used by children when they are still not able to express themselves by means of words. Following Papaeliou & Trevarthen (2006), we hypothesize that communicative vocalizations would be shorter than investigative ones. We further suggest that when producing communicative vocalizations children would use prosodic and gestural cues in order to express specific pragmatic meanings such as discontent, requests, responses to stimuli, or general statements. In fact, our suggestion is that language does not start when children produce their first words, but that it is already functioning by means of the communicative usage of gestures and prosody.

2. **Experimental study 2.** Children synchronize gesture and speech before than what is suggested. Previous studies suggest that children do not integrate gesture and speech in an adult-like way until the two-word stage (Butcher & Goldin-Meadow 2000). We hypothesize that children produce co-speech gestures and synchronize gesture and speech in an adult-like manner before the two-word stage. This target alignment for children would be one in which the pitch peak and pitch accent of the speech is produced during the stroke of the gesture (McNeill, 1992; Kendon, 1980).

3. **Experimental study 3.** In case that the first and second hypotheses are confirmed, and given that in language development, most language abilities are first understood than produce, we assume that children understand communicative intentions before they can produce them. Plus, we hypothesize that children are able to understand other’s communicative intentions relying on gestural and prosodic communicative cues.

These three experimental studies will show whether babbling infants use prosody to produce and comprehend communicative vocalizations with intentionality, and how prosody and gesture work together to entail the intended meaning. The dissertation will be organized as follows. First, an introductory chapter with theoretical framework and
previous work in the field will be presented and discussed. Second, the goals of the study and the hypotheses will be exposed. Then, three experimental studies to test the hypotheses will be described, with details about materials, methods, results, conclusion, and discussion. Finally, the conclusions that may be drawn from the three experimental studies will be laid out.

5. EXPERIMENTAL STUDIES

5.1. Study 1

5.1.1. Materials

The first study aims at investigating if gesture and prosody are the communicative cues used by children when they are still not able to express themselves by means of words. In order to do it, a longitudinal corpus of four Catalan-monolingual children will be used to study if gesture and prosody are the communicative cues used by babbling children to communicate.

This corpus, Esteve-Prieto corpus, is an online Phon database of four monolingual Catalan speaking-children that were recorded from 0;6 to 2;0 (still in progress): An, Bi, Ma, and On. We plan to incorporate this corpus in the CHILDES database, once it will be phonetically and orthographically transcribed. It is available in the following website: [http://prosodia.upf.edu/phon/en/corpora/description/esteveprieto.html]

All four children were video-recorded at their homes during weekly 30-minute sessions using a SONY camera, model DCR-DVD202E PAL. Recordings were made by the author of this Ph.D. project, who was previously acquainted with the families and children. Children were always recorded in the same room of their respective homes, typically in their living-rooms, during free play sessions. All children were recorded as they interacted with their mothers except for one child, An, who was recorded while interacting with both her father and her mother in most of the sessions. A tripod was used, placed as close to the child as possible and positioned so that the camera was pointing toward the child’s face.
We selected for analysis vocalizations produced at three specific points in time, i.e. when children were 0;7, 0;9, and 0;11. Table 1 summarizes the data for all sessions included in this study. These ages were selected based on the hypothesis that these vocalizations would display the typical features of certain stages of development: before the onset of intentional communication, when intentionality starts, and when intentionality is already developed (Piaget, 1936, 1946; Trevarthen, 1977, 1979, 1982, among others).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Duration</th>
<th>Participant</th>
<th>Age</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>An</td>
<td>0;7.10</td>
<td>0:30:09</td>
<td>Bi</td>
<td>0;7.02</td>
<td>0:31:07</td>
</tr>
<tr>
<td></td>
<td>0;7.17</td>
<td>0:28:22</td>
<td></td>
<td>0;7.09</td>
<td>0:32:02</td>
</tr>
<tr>
<td></td>
<td>0;7.29</td>
<td>0:34:03</td>
<td></td>
<td>0;7.16</td>
<td>0:30:33</td>
</tr>
<tr>
<td></td>
<td>0;9.04</td>
<td>0:36:39</td>
<td></td>
<td>0;7.22</td>
<td>0:33:24</td>
</tr>
<tr>
<td></td>
<td>0;9.11</td>
<td>0:19:03</td>
<td></td>
<td>0;9.23</td>
<td>0:30:36</td>
</tr>
<tr>
<td></td>
<td>0;9.18</td>
<td>0:36:45</td>
<td></td>
<td>0;9.29</td>
<td>0:33:05</td>
</tr>
<tr>
<td></td>
<td>0;9.25</td>
<td>0:37:30</td>
<td></td>
<td>0;11.05</td>
<td>0:34:41</td>
</tr>
<tr>
<td></td>
<td>0;11.03</td>
<td>0:33:00</td>
<td></td>
<td>0;11.12</td>
<td>0:36:20</td>
</tr>
<tr>
<td></td>
<td>0;11.08</td>
<td>0:36:34</td>
<td></td>
<td>0;11.18</td>
<td>0:34:21</td>
</tr>
<tr>
<td></td>
<td>0;11.15</td>
<td>0:36:35</td>
<td></td>
<td>0;11.25</td>
<td>0:26:09</td>
</tr>
<tr>
<td></td>
<td>0;11.22</td>
<td>0:33:20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ma</td>
<td>0;7.06</td>
<td>0:36:58</td>
<td>On</td>
<td>0;7.02</td>
<td>0:38:01</td>
</tr>
<tr>
<td></td>
<td>0;7.13</td>
<td>0:27:19</td>
<td></td>
<td>0;7.15</td>
<td>0:34:29</td>
</tr>
<tr>
<td></td>
<td>0;7.20</td>
<td>0:26:33</td>
<td></td>
<td>0;7.22</td>
<td>0:33:10</td>
</tr>
<tr>
<td></td>
<td>0;9.04</td>
<td>0:19:03</td>
<td></td>
<td>0;9.03</td>
<td>0:36:08</td>
</tr>
<tr>
<td></td>
<td>0;9.09</td>
<td>0:34:02</td>
<td></td>
<td>0;9.11</td>
<td>0:34:40</td>
</tr>
<tr>
<td></td>
<td>0;9.17</td>
<td>0:37:15</td>
<td></td>
<td>0;9.16</td>
<td>0:36:30</td>
</tr>
<tr>
<td></td>
<td>0;9.29</td>
<td>0:35:00</td>
<td></td>
<td>0;11.14</td>
<td>0:26:25</td>
</tr>
<tr>
<td></td>
<td>0;11.00</td>
<td>0:32:29</td>
<td></td>
<td>0;11.23</td>
<td>0:37:12</td>
</tr>
<tr>
<td></td>
<td>0;11.05</td>
<td>0:34:43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;11.12</td>
<td>0:39:44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;11.19</td>
<td>0:35:20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;11.25</td>
<td>0:33:23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TOTAL AMOUNT OF TIME RECORDED:** 23:02:10

Table 1. Recorded sessions included in the study, classified by children’s age, and duration of the session.

5.1.2. Codification

The approximately 20 hours of recordings were segmented into more than 3,000 vocalizations. From these, around 300 were excluded from the analysis because of the following circumstances: (1) when child and parent overlapped when vocalizing, (2) when ambient noise was too loud, (3) when the child vocalized while having an object
inside his/her mouth, and (4) when the sound did not show a visible trace on the spectrogram.

After segmenting vocalizations, three main analyses were conducted with data: an acoustic analysis, a gestural analysis, and a pragmatic analysis. We explain now in more detail the specificities of these analyses.

First, the pragmatic analysis was performed using Phon software system (Rose, McWhinney, Byrne, Hedlund, Maddocks, O'Brien, & Warehem, 2005). All vocalizations were first annotated in terms of the communicative function they conveyed. Since the one of the aims was to discover whether the vocalizations conveying communicative information are different from when not intending to communicate, we first classified them as being ‘communicative’ or ‘investigative’. Following Papaeliou & Trevarthen (2006), a vocalization was considered to be investigative if the infant was holding an object, inspecting an object or completing a task; a vocalization was considered to be communicative if the child was interacting with an adult, pointing, directing eye-gaze to the adult, and reaching or giving something. Thus, communication was established in a context of joint attention where triadic relations between the parent, the child, and an object or event were observed. Since we also hypothesized that children use prosody and gestures to express specific pragmatic meanings, a more specific pragmatic analysis was conducted. Thus, all communicative vocalizations were further divided into these categories:

- Request (speech act)
- Response to a stimulus (speech act)
- General statement (speech act)
- Vocative (speech act)
- Discontent (emotion)
- Satisfaction (emotion)
- Surprise (emotion)

These categories could overlap. Hence, it could be the case that a communicative vocalization was then annotated as a ‘request’ with ‘discontent’, since the child could require the other to do something by means of showing discontent towards the actual state of facts.
Second, the gestural analysis was performed in parallel with the pragmatic analysis explained above. Each vocalization was annotated in terms of the gestures displayed by children when vocalizing, using Phon software system (Rose et al., 2005). Table 2 shows the gesture categories used (based on Allwood, Cerrato, Jokinen, Navarretta, & Paggio, 2007).

<table>
<thead>
<tr>
<th>Gaze direction</th>
<th>absent gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gaze to camera</td>
</tr>
<tr>
<td></td>
<td>gaze to object</td>
</tr>
<tr>
<td></td>
<td>gaze to parent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Manual gestures</th>
<th>clapping hands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>extending arms</td>
</tr>
<tr>
<td></td>
<td>embracing parent</td>
</tr>
<tr>
<td></td>
<td>manipulating object</td>
</tr>
<tr>
<td></td>
<td>moving arms</td>
</tr>
<tr>
<td></td>
<td>pointing object</td>
</tr>
<tr>
<td></td>
<td>moving hands</td>
</tr>
<tr>
<td></td>
<td>shaking arms</td>
</tr>
<tr>
<td></td>
<td>no specific manual gesture</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facial gestures</th>
<th>furrowing brows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>opening eyes</td>
</tr>
<tr>
<td></td>
<td>closing eyes</td>
</tr>
<tr>
<td></td>
<td>opening mouth</td>
</tr>
<tr>
<td></td>
<td>closing mouth</td>
</tr>
<tr>
<td></td>
<td>pouting</td>
</tr>
<tr>
<td></td>
<td>shaking head</td>
</tr>
<tr>
<td></td>
<td>smiling</td>
</tr>
<tr>
<td></td>
<td>rising eyebrows</td>
</tr>
<tr>
<td></td>
<td>no specific facial gesture</td>
</tr>
</tbody>
</table>

Table 2. Gesture categories used in the gesture analysis.

Third, the acoustic analysis was performed to find out whether different prosodic patterns are at play when infants try to communicate or convey a set of pragmatic functions. In order to perform the acoustic analysis, we manually extracted all the audio files (in the .wav format) from our Phon corpus and analyzed them with the Praat software package (Boersma & Weenink, 2005). A set of prosodic features were manually labeled: duration and pitch contour, i.e. pitch maximum and minimum points, and starting and end points of vocalizations. The aim was to analyze global pitch range of the contour and total duration, which are the features that are most commonly used in studies of the prosody of infants' vocalizations (Marcos, 1987; Papaeliou et al., 2002; Papaeliou & Trevarthen, 2006; Scherer, 1986).
On the one hand, to obtain the pitch range value, we selected three pitch points from the fundamental frequency contour: the first pitch point (p1) representing the pitch value at the beginning of the vocalization, also referred to as the reference level of the speaker; the second pitch point (p2) selected at the maximum peak in the fundamental frequency contour; and finally, the third point (p3) representing the pitch value at the end of the vocalization. The pitch range was then calculated by subtracting the minimum pitch value from the maximum pitch. In order to compare different pitch ranges across the three children, the pitch values were extracted in semitones and in Hz. On the other hand, to obtain the duration value, in order to obtain the total duration of the vocalization, the first point (t1) and last point (t2) in the fundamental frequency line were selected. Following Papaeliou & Trevarthen’s (2006) work, we considered two sounds to be distinct vocalizations if they were separated by at least 50 ms. Figure 1 shows an example of the acoustic annotation in Praat.

![Figure 1. Illustration of a vocalization performed by Ma, at 0;11.](image)

5.1.3. Summary results

Statistical analysis of the data showed that pitch range and duration are both significantly affected by the communicative status of the vocalization. As for pitch range, vocalizations display wider pitch range when the children are communicating than when they are performing investigative vocalizations. Further analyses of communicative vocalizations have revealed that depending on the pragmatic intention expressed, pitch range is wider or narrower. For instance, protests and requests have wider pitch ranges than responses and statements.
As for duration, communicative vocalizations are shorter in general than investigative ones. Yet these results also seem to show that the duration cue is not controlled until children are 0:9. Our subsequent analysis of communicative vocalizations, whereby they were separated into specific pragmatic intentions, showed that the durational patterns of the vocalizations are strongly influenced by pragmatic functions. Specifically, responses and statements behave similarly, being shorter than the other pragmatic functions. Protests are the longest vocalizations, followed by requests, and
in the middle there are satisfactions, which are shorter than protests and requests but longer than responses and statements.

![Figure 4](image1.png)

*Figure 4. Error bars of the duration of vocalizations for all infants, separated by age and communicative status.*

![Figure 5](image2.png)

*Figure 5. Box plots of the duration of vocalizations at three different ages for all infants, broken down by pragmatic function.*

The analysis of gestural patterns (which included annotation of gaze direction, manual gestures, and facial gestures produced by children) revealed a very close relationship between gestural cues and communicativeness, as well as between gestural cues and intentionality. Results suggested that children’s gestures are clearly correlated with the classification of a certain vocalization as communicative (for instance, pointing or furrowing of the brows).
5.2. Study 2

5.2.1. Materials

The second study aims at investigating how children compared to adults phonologically synchronize gesture and speech, and if they do it before what is suggested in previous studies. Thus, two experiments will be carried out. First, an experiment is to see if adult Catalan-speakers temporally align gesture and speech in a way that the pitch peak and the pitch accent of the speech is produced during the stroke of the gesture. Second, an experiment is carried out to see if children temporally align gesture and speech in an adult-like manner and when they start doing so.

**FIRST EXPERIMENT: TEMPORAL ALIGNMENT IN ADULTS**

For the first experiment, twenty adult Central Catalan-speakers will be recorded while performing a pointing-naming task. For the recordings, a Panasonic HD AVCCAM will be used. Following Rochet-Capellan et al. (2008) procedure, participants seat in a chair approximately 50 cm far from the screen. In the screen, a target (smiley symbol) and a word will be projected at the same time (see figure 6). Participants perform a pointing-naming task in which they have to point at the smiley face while reading the word at the precise moment when the smiley face turned from red into green color. The target words differ in their number of syllables and stress position: /pá/, /papá/, /pápa/, /tá/, /tatá/, /táta/, /má/, /mamá/, /máma/, /ná/, /naná/, and /nána/.  

![Figure 6: Example of an image projected in the screen during the pointing-naming task.](image.png)

**SECOND EXPERIMENT: TEMPORAL ALIGNMENT IN CHILDREN**
For the second experiment, recordings from the Esteve-Prieto corpus will be used. This corpus is an online Phon database of four monolingual Catalan speaking-children that were recorded from 0;6 to 2;0 (still in process): An, Bi, Ma, and On (for more information, see section 5.1.1). We selected for the analysis communicative acts produced by these children at 0;11, 1;1, 1;3, 1;5, and 1;7. Vocal speech acts, gestural speech acts, and combinations of the two modalities were all included in the analysis. These ages were selected in the basis of the word production. At 0;11, children were only producing routine words such as mama or papa, so it could be stated that it was before the one-word period. At ages 1;1, 1;3, all children started producing target words regularly, and at 1;5 they were already at the one-word stage. At 1;7, some children started producing their two-word combinations, such as mama aigua (‘mummy water’) or mama ja’tà (‘mummy I am done’). Therefore, by selecting this age range we will have an overview form the babbling stage until the beginning of the two-word period. Table 3 summarizes the data for all sessions included in this study.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Duration</th>
<th>Participant</th>
<th>Age</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>An</td>
<td>0;11.03</td>
<td>0:33:00</td>
<td>Bi</td>
<td>0;11.05</td>
<td>0:34:41</td>
</tr>
<tr>
<td></td>
<td>0;11.08</td>
<td>0:36:34</td>
<td></td>
<td>0;11.12</td>
<td>0:36:20</td>
</tr>
<tr>
<td></td>
<td>0;11.15</td>
<td>0:36:35</td>
<td></td>
<td>0;11.18</td>
<td>0:34:21</td>
</tr>
<tr>
<td></td>
<td>0;11.22</td>
<td>0:33:20</td>
<td></td>
<td>0;11.25</td>
<td>0:26:09</td>
</tr>
<tr>
<td></td>
<td>0;12.10</td>
<td>0:37:21</td>
<td></td>
<td>0;12.20</td>
<td>0:34:05</td>
</tr>
<tr>
<td></td>
<td>0;12.24</td>
<td>0:41:48</td>
<td></td>
<td>0;13.07</td>
<td>0:34:59</td>
</tr>
<tr>
<td></td>
<td>0;15.07</td>
<td>0:29:10</td>
<td></td>
<td>0;15.15</td>
<td>0:35:57</td>
</tr>
<tr>
<td></td>
<td>0;15.28</td>
<td>0:34:49</td>
<td></td>
<td>0;15.29</td>
<td>0:35:31</td>
</tr>
<tr>
<td></td>
<td>0;17.20</td>
<td>0:25:29</td>
<td></td>
<td>0;17.03</td>
<td>0:37:55</td>
</tr>
<tr>
<td></td>
<td>0;17.29</td>
<td>0:34:05</td>
<td></td>
<td>0;17.17</td>
<td>0:37:58</td>
</tr>
<tr>
<td></td>
<td>0;19.05</td>
<td>0:35:42</td>
<td></td>
<td>0;19.26</td>
<td>0:37:12</td>
</tr>
<tr>
<td></td>
<td>0;19.16</td>
<td>0:34:21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ma</td>
<td>0;11.05</td>
<td>0:34:43</td>
<td></td>
<td>0;11.14</td>
<td>0:26:25</td>
</tr>
<tr>
<td></td>
<td>0;11.12</td>
<td>0:39:44</td>
<td></td>
<td>0;11.23</td>
<td>0:37:12</td>
</tr>
<tr>
<td></td>
<td>0;11.19</td>
<td>0:35:20</td>
<td></td>
<td>0;11.28</td>
<td>0:36:15</td>
</tr>
<tr>
<td></td>
<td>0;11.25</td>
<td>0:33:23</td>
<td></td>
<td>0;15.08</td>
<td>0:23:28</td>
</tr>
<tr>
<td></td>
<td>0;12.20</td>
<td>0:32:46</td>
<td></td>
<td>0;15.21</td>
<td>0:36:43</td>
</tr>
<tr>
<td></td>
<td>0;13.14</td>
<td>0:31:17</td>
<td></td>
<td>0;17.15</td>
<td>0:37:09</td>
</tr>
<tr>
<td></td>
<td>0;13.27</td>
<td>0:33:36</td>
<td></td>
<td>0;19.14</td>
<td>1:10:54</td>
</tr>
<tr>
<td></td>
<td>0;15.08</td>
<td>0:35:48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;15.22</td>
<td>0:32:56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;17.23</td>
<td>0:34:51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;19.05</td>
<td>0:36:29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On</td>
<td>0;11.05</td>
<td>0:34:43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;11.12</td>
<td>0:39:44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;11.19</td>
<td>0:35:20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;11.25</td>
<td>0:33:23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;12.20</td>
<td>0:32:46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;13.14</td>
<td>0:31:17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;13.27</td>
<td>0:33:36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;15.08</td>
<td>0:35:48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;15.22</td>
<td>0:32:56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;17.23</td>
<td>0:34:51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0;19.05</td>
<td>0:36:29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2.2. Codification

**FIRST EXPERIMENT: TEMPORAL ALIGNMENT IN ADULTS**

For the first experiment with adults, all pointing-naming instances produced by the 20 adult Catalan-speakers will be acoustically and gesturally analyzed using ELAN (Lausberg & Sloetjes, 2009) and Praat software package (Boersma & Weenink, 2005). The aim is to investigate how Catalan speakers synchronize gesture and speech in their productions. Previous research has found that the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable in speech (Kendon, 1980; McNeill, 1992; Rochet-Capellan et al., 2008). With the present analysis we will investigate if the same occurs in Catalan and which are the values of the precise alignment of stroke and pitch contour. Thus, two analyses will be carried out: an acoustic analysis and a gestural analysis.

For the gestural analysis, all pointing gestures will be annotated in terms of gestural phases. Following McNeill (1992) the phases of deictic gestures are:

- Preparation (optional in the gesture)
- Pre-stroke hold (optional in the gesture)
- Stroke (the only compulsory one)
- Post-stroke hold (optional in the gesture)
- Retraction (optional in the gesture)

Figure 7 shows an example of these phases in an adult pointing gesture.
Figure 7. Gesture phases and the position of the pitch peak, intensity peak, or pitch accent in respect to the gesture phases when the gesture-speech combination is adequately aligned.

For the acoustic analysis, three values will be annotated using Praat (see figure 8b):
- The pitch peak point in the fundamental frequency line
- The pitch accent
- The intensity peak point

Afterwards, the position of these three acoustic values will be imported in ELAN. Figure 8a shows an example of annotation of the pointing-naming task in ELAN. In this figure, the top left panel displays the video images, the top right panel displays the controls, and the bottom panel includes the annotation tiers where prosodic and gestural cues are annotated: type of gesture, gesture phases, word uttered, position of the pitch accent, and position of the pitch peak point.
SECOND EXPERIMENT: TEMPORAL ALIGNMENT IN CHILDREN

For the second experiment with children, all communicative acts produced by children at 0;11, 1;1, 1;3, 1;5, and 1;7 were classified as ‘speech-only’, ‘gesture-only’, or ‘gesture-speech’ combinations. Then, ‘gesture-speech’ combinations were analyzed in terms of prosody and gestures using ELAN and Praat software package. The aim of the analysis is to investigate the temporal synchronization between gesture and
speech, i.e. whether the gestural stroke was aligned with the pitch peak in vocalizations.

As for the gestural analysis, two values are annotated using Elan: type of gesture and gesture phases. First, the type of gesture is codified. Relying on the visual cues, each communicative act involving a gesture had to be annotated as containing a ritualized gesture (saying bye-bye, shaking head to negate, extending the finger to signal one year), a deictic gesture (extending his/her arm, pointing towards an object or an event), or an iconic gesture (for instance, with the hand by the ear to signal telephone).

Second, the gesture phases of deictic gestures were further codified in order to analyze its temporal synchronization with speech. Figure 9 shows two examples of the gesture phases in a pointing gesture. Following McNeill (1992), the gesture phases used were:

- Preparation (optional in the gesture)
- Pre-stroke hold (optional in the gesture)
- Stroke (the only compulsory one)
- Post-stroke hold (optional in the gesture)
- Retraction (optional in the gesture)

Figure 9. Examples of gesture phases in a deictic gesture
For the prosodic analysis, the precise location of the pitch peak, i.e. the highest point in the fundamental frequency line, was annotated using Praat (see figure 10b). Thus, a ‘gesture-speech’ combination was considered to be adequately aligned when the pitch peak of the fundamental frequency line occurred during the stroke of the gesture.

Figure 10a is an example of annotation in ELAN. In the top left of the ELAN window, the video images are displayed; on the top right, controls such as speed, volume, or navigation through the cases annotated can be controlled; on the bottom panel of the window, all gestural and prosodic cues are annotation in different tiers: type of gesture, gesture phases, verbal speech acts, position of the pitch peak, and gesture intention. Besides, the waveform is displayed between the top and the bottom panels, in order to export audio to Praat and to allow annotation of the prosodic information such as pitch range.

Figure 10a. Example of annotation of the children’s data in ELAN. In this example, the pitch peak of the speech (in the bottom green circle) occurs after the stroke of the pointing gesture (in the top red circle).

Figure 10b. Annotation of the pitch peak in Praat. When this point is located, its position is imported in Elan, as shown in the green circle in figure 10a.
5.2.3. Preliminary results

FIRST EXPERIMENT: TEMPORAL ALIGNMENT IN ADULTS

Preliminary results after analyzing only five adult Catalan-speakers show that speakers temporally align the stroke of the pointing gestures with the pitch accent in 98% of the cases (see figure 11). No significant differences were found between monosyllabic and disyllabic words, and neither on the position of the pitch accent in disyllabic words.

![Figure 11. Percentage of synchronized and non-synchronized gesture-speech combinations.](image)

When analyzing the position of the pitch peak in respect of the stroke of the gesture, results showed that the pitch peak tended to align with the beginning of the stroke (see figure 12).
SECOND EXPERIMENT: TEMPORAL ALIGNMENT IN CHILDREN

Results showed that at 0;11, from all communicative acts containing gesture, only 43.26% of them were 'gesture-speech' combinations. However, already at 1;1 the 'gesture-speech' combinations represent 54.8% of the communicative acts containing gesture, and children produce even higher proportions of 'gesture-speech' combinations in the late babbling and one-word stage (see Figure 13).
The analysis of the temporal alignment of the ‘gesture-speech’ combinations and, specifically, the deictic gestures accompanied by speech, showed that at 0;11, children still do not synchronize gesture and speech, since they produce the pitch peak during the stroke of the gesture only in 30% of the cases. At this early age, children produce most of the pitch peaks either before or after the stroke of the gesture. However, at the late babbling stage and early one-word period, results suggest that children already produce a higher proportion of synchronized ‘gesture-speech’ combinations than unsynchronized ones (see figure 14).

![Figure 14. Evolution of the position of the pitch peak compared to the stroke. There is synchronization when the pitch peak is produced during the stroke.](image)

A closer analysis of the temporal alignment showed that at the beginning of the babbling stage, the pitch peak tends to be aligned at the end of the stroke of the deictic gesture. However, at the late babbling stage and the one-word period, the pitch peak has moved to the left and it is aligned at the beginning of the stroke (see figure 15), just like adults seemed to do in the first experiment of this study.
Figure 15: Longitudinal evolution of the alignment of the pitch peak and the stroke. At the early babbling stage, the pitch peak is aligned at the end of the stroke, but at the one-word period it is aligned at the beginning of the stroke.

5.3. Study 3

5.3.1. Materials

The third study aims at investigating if children are able to understand other’s communicative intentions relying on gestural and prosodic communicative cues. It will be a study in collaboration with a research group will be carried out. This research group, called Communication Before Language Group, is based in The Max Planck Institute for Psycholinguistics and coordinated by Prof. Ulf Liszkowski. Their research focuses on infants' gestural communication, prelinguistic infants' social-interactional experiences across different cultures, infants' understanding of mental states, and infants' expectations of and motivation for joint collaborative activities.
Interestingly, after a first meeting with our collaborators, a proposal is starting to emerge: the cross-linguistic comparison would give us stronger evidence of this children’s capacity. Thus, languages like Catalan, Spanish, or Dutch use different prosodic contours to encode requests and informative utterances. Whenever these differences arise in the children’s ability to understand adults’ intentions, it would be clear that they rely on prosodic cues to understand the different intentions behind a pointing gesture.

5.3.2. Codification

The hypothesis of the third study will be tested in collaboration with the research group Communication Before Language Group (Max Planck Institute for Psycholinguistics, Nijmegen). Details on this project still have to be defined. However, the first contact has already been established, and a proposal has already been formulated.

By means of an eye-tracking system, a methodology used to record eye movements, we will study the ability to understand adult’s intentions by babbling children. Since the children we are planning to test are still not fluent in language production, the best methodology will be one that does not rely on the verbalization of children but only on how they cognitively process a certain situation.

In a first version of the proposal, a draft of the methodology to be used has been agreed. The study aims at investigating if the parents’ use of specific request prosody (in Catalan we have a specific contour for requests, a L* HL% or even a L* LHL% if it is more insistent) when producing a pointing gesture increases more clear responses on the part of children in taking the objects. The prediction would be that when the experimenter uses request intonation types, the child will respond better to the request than when using the neutral intonation. And even more when using the insistent request intonation.

For Catalan, different intonation contours could be used:
- With a neutral intonation (H*L%)
- With a requestive intonation (L* HL%)
- With an insistent requestive intonation (H* LHL%)
Thus, we predict that when the experimenter uses a requestive intonation, the babbling child will respond better to the request than when using the neutral intonation. And this will happen even more often when using the insistent requestive intonation.

### 6. WORKING SCHEDULE

| June 2011                  | ° Submission of the Ph.D. project.  
|                           | ° Oral presentation of results of the second part of study 2 (alignment in children) in the conference *Phonetics and Phonology in Iberia (PaPI)*, held in Tarragona. |
| July 2011 to October 2011  | ° Correction of an article submitted in the *Journal of Child Language* with results from study 1 (gesture and prosody to communicate intentionally), according to the revision proposed by the journal.  
|                           | ° Presentation of results of the second part of study 2 (alignment in children) in the conference AMLaP (Architectures and Mechanisms for Language Processing), to be held in Paris, France.  
<p>|                           | ° Presentation of results of the second part of study 2 (alignment in children) in the conference GESPIN (Gesture and Speech in Interaction), to be held in Bielefeld, Germany. |
| November 2011 to January 2012 | ° Writing the results of the second part of study 2 (alignment in children) with results and comments obtained from the conferences. |
| February 2012 to March 2012 | ° Gathering the complete sample of first part of study 2 (alignment in adults). |
| April 2012 to Mai 2012     | ° Writing the results of the complete study 2 (alignment in adults and alignment in children). |</p>
<table>
<thead>
<tr>
<th>Dates</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 2012 to July 2012</td>
<td>Preparation of materials for study 3 (children’s understanding of intentions in pointing).</td>
</tr>
<tr>
<td>September 2012 to December 2012</td>
<td>Research stay in the Communication Before Language Group (Max Planck Institute for Psycholinguistics) to collaborate in the study 3.</td>
</tr>
<tr>
<td>January 2013 to April 2013</td>
<td>Analysis of results from the study 3.</td>
</tr>
<tr>
<td></td>
<td>Writing study 3.</td>
</tr>
<tr>
<td>Mai 2013 to July 2013</td>
<td>Writing the Ph.D. dissertation.</td>
</tr>
<tr>
<td>September 2013</td>
<td>Dissertation defense.</td>
</tr>
</tbody>
</table>

Butcher & Goldin-Meadow (2000) analyzed 6 English-speaking children from the one-word stage to the two-word stage in order to see (1) if communicative gestures form an integrated system with speech in the one-word stage, (2) if there is a time early in development when communicative gesture is used primarily without speech, and (3) if young children do produce gesture in combination with speech, are the two modalities integrated both temporally and semantically, as they are in adults. Their results suggest that at some point during the one-word period, children begin to combine gesture and speech in an adult-like way, and at the same moment produce those gestures in temporal and semantic synchrony with that speech.


Papaeliou & Trevarthen (2006) observed four English-speaking infants from 0;7 to 0;11 and classified their vocalizations as ‘communicative’ or ‘investigative’ according to concurrent non-vocal behaviors. They found that children displayed different prosodic patterns when vocalizations were classified as communicative, or when classified as investigative: compared to investigative vocalizations, communicative vocalizations had a higher mean and maximum f(0), higher standard deviation of f(0), and shorter duration. Hence, the authors concluded that pre-linguistic vocalizations can be a tool for both communicating and thinking.


Prieto et al. (in press) investigated the development of prosodic patterns in four Catalan-learning children and in two Spanish-learning children between 0;11 and 2;4.
Their results show that (1) the Autosegmental Metrical Model was successful in transcribing early intonation contours, (2) that children’s emerging intonation is independent of grammatical development and that it develops before the two-word combinations, and (3) that the emergence of intonational grammar is related to the onset of speech and the presence of a small lexicon, since they found that children at 1;1 and 1;3 are able to produce a set of adult-like intonation contours.


Rochet-Capellan et al (2008) analyzed 20 Portuguese Brazilian-speakers to analyze the effects of stress on jaw-finger coordination. To do it, participants had to point to a target while naming it using different target words with a trochee or iamb pattern. Results found a strong synchronization between arm and jaw movements, since the pitch accent occurred simultaneously with the stroke of the gesture, both in the case of trochees and in the case of iambs. According to the authors, these results evidence that the speech deictic site (namely, the part of speech that shows) is anchored in the pointing gesture.


Tomasello et al. (2007) propose a theory in which infant pointing is the evidence of the infants’ early intentionality and shared intentionality. They propose three different motives behind the infant pointing: (1) the infant want the adult to do something, (2) the infant want the adult to know something, and (3) the infant want the adult to feel something. The authors review many articles finding results that support the hypothesis that infant pointing gestures reflect their motivation for cooperation and shared intentionality, concluding that at 1;0-1;2, children build their linguistic skills on the basis of this prelinguistic communication.
8. GENERAL REFERENCES


