The integration of prosody and gesture in early intentional communication

Núria Esteve Gibert

TESI DOCTORAL UPF / 2014

DIRECTOR DE LA TESI

Dra. Pilar Prieto

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL LLENGUATGE

Universitat Pompeu Fabra Barcelona

A la meva família

Acknowledgments

I have learned many things by doing this dissertation, and I am indebted to many people for that. First, I am indebted to my advisor Pilar Prieto because she has offered me her unconditional time, support and knowledge. She was always available when needed, always happy to help. Thanks for taking my ideas seriously, for trusting in me, for opening my mind, and for your great guidance over the last four and a half years. It has been a great pleasure to work with you and I will always be grateful for that.

I am also grateful to Ulf Liskowzki for hosting me at the Max Planck Institute for Psycholinguistics in Nijmegen. You agreed to host me without knowing me at all, and this has taught me a lot. With you and your team (Mundi, Marloes, Mireille, and Margret) I have learned about pointing gestures, about how a Babylab works, about how to design a experiment with infants, and about how research is done in a leading institution. I also want to thank Ferran Pons and Laura Bosch because they have always helped me when I needed it, encouraged me in my research, and opened the doors of their lab to me. Thank you very much.

Additional thanks are owed to the Department of Translation and Language Sciences at the Universitat Pompeu Fabra. I give special thanks to the Head of Department Àlex Alsina and to Toni Badia, Aurora Bel, Carmen Pérez, and Louise McNally, for their support during my pre-doctoral stage. Thanks for your funding, for the grant to do the three-month research stay, for the financial support that allowed me to go to many conferences where I learned about some

cutting-edge research, and for bringing to Barcelona some of the top researchers in the world. Also, thanks for giving me the chance to discover that I actually like teaching a great deal.

I would also like to thank various members of academia from whom I have learned in one way or another. Thank you Llorenç Andreu, Meghan Armstrong, Amalia Arvaniti, Luca Bonatti, Aoju Chen, Yiya Chen, Judith Holler, José Ignacio Hualde, Paolo Roseano, Marc Swerts, and Maria del Mar Vanrell.

I truly thank all members of the Group of Prosodic Studies, namely Alfonso, Iris, Joan, Maria del Mar, Meghan, Paolo, Rafèu, Santi, and Vero, and my other officemates from the Linguistics Research Unit at the Universitat Pompeu Fabra. Thanks for showing me what a great research group is and how a real team functions. But especially thanks for sharing conversations, lunches, coffees, more coffees, some beers, songs, hostel rooms, and sunbaths at the Plaça Gutenberg in Barcelona.

Big, big thanks to the parents of the children I recorded from when they started goo-gooing at 6 months of age until they were on the verge of mastering language at 3 years of age. Rosalba and Jose, Txell and Òscar, Ainhoa and Guillem, Sònia and Toni: you were always happy to receive me, and without your disinterested help I would never have been able to carry out two of the studies described in this dissertation. And thanks to their children themselves, Àngela, Biel, Martí, and Ona, for giving me the chance to experience their first pointing and first word. With you I have seen in real time the process of language development, and that is

totally priceless. And thanks to all the other participants in my studies, because I don't like asking for favors and I have had to do it a lot over the last few years.

And special thanks to all the people who are close to me. I thank all my friends and especially the great *Teresines* Lidi, Pili, and Olga, and also Ariadna, Joan Terrones, and Ju. Talking to you and laughing with you is like coming back to life. You know that while I was sitting here in front of the computer you guys partied, got married, and had children. But now it's my turn (haha, just joking!).

I will never be able to express sufficient gratitude to dear my family: my amazing parents, Núria and Francesc, who have so generously worked hard for their family and from whom I have learned everything important I know. Thanks for respecting me and having such confidence in me. I also thank my sisters Anna and Montse, two wonderful women with wonderful families, essential supports for me. And my *tieta* Montse, a fantastic woman with whom I have learned a lot and shared great moments. And my wonderful *iaia* Maria, all wisdom and energy. You are an example for me of how to live life. And last but not least, immense thanks to Jaume for sharing your life with me, for your incredible generosity, for helping me to remember what is important in life. I would not have been able to do it without you.

Abstract

This dissertation comprises four experimental studies which investigate the way infants integrate prosody and gesture for intentional communicative purposes. As adult speakers we automatically integrate prosody and gestures at a temporal and pragmatic level, and we use these cues together with social contextual information to convey and understand intentional meanings. My aim is to investigate whether infants use prosodic and gesture features in an integrated way for communicative purposes prior to their use of lexical-semantic cues. The dissertation includes four studies, each one described in a separate chapter. The first study is a longitudinal analysis of how a group of infants produce gesture and speech combinations in natural interactions, with results that show that already at 12 and 15 months of age infants temporally align prosodic and gesture prominences. The second study uses a habituation/test procedure to test the infants' early sensitivity to temporal gesture-prosodic integration, showing that 9-month-old infants are sensitive to the alignment between prosodic and gesture prominences. The third study analyzes the longitudinal productions of four infants at the pre-lexical stage and provides evidence that infants use prosodic cues such as pitch range and duration to convey specific intentions like requests, statements, responses, and expressions of satisfaction or discontent. Finally, the fourth study examines how infants responded at 12 months of age to different types of pointing-speech combinations and shows that

infants use prosodic and gestural cues to comprehend communicative intentions behind an attention-directing act. Altogether, this dissertation shows that the temporal integration of gesture and speech occurs at the early stages of language and cognitive development, and that pragmatic uses of prosody and gesture develop before infants master the use of lexical cues. Thus, prosody is the first grammatical component of language that infants use for communicative purposes, revealing that linguistic communication emerges before infants have the ability to use lexical items with semantic meanings. I further claim that infants' integration of prosody and gesture at the temporal and pragmatic levels is a reflex of an early emergence of language pragmatics.

Resum

Aquesta tesi inclou quatre estudis experimentals que investiguen com els infants integren prosòdia i gestualitat amb fins comunicatius. Els adults integrem la prosòdia i la gestualitat de manera temporal i pragmàtica i, juntament amb la informació sociocontextual, ho utilitzem per a transmetre i comprendre significats intencionals. En aquesta tesi es pretén investigar si els infants utilitzen la prosòdia i la gestualitat de manera integrada per a fins comunicatius, abans de ser capaços d'emprar elements lexicosemàntics. La tesi inclou quatre estudis, cada un en un apartat diferent. El primer estudi analitza longitudinalment les combinacions de gest i parla dels infants en interaccions espontànies, i mostra que a partir dels 12 o 15 mesos els infants alineen temporalment la prominència prosòdica i la prominència gestual. El segon estudi empra el mètode d'habituació/test per a comprovar l'habilitat primerenca dels infants a percebre la integració temporal entre prosòdia i gest, i mostra que als 9 mesos els infants ja són capaços de percebre l'alineació temporal entre la prominència gestual i la prosòdica. El tercer estudi també analitza longitudinalment les produccions dels infants en interaccions espontànies per a mostrar que, abans de produir les primeres paraules, els infants ja utilitzen elements prosòdics com el rang tonal i la durada, a més del gest, per a transmetre actes de parla com ara la petició, les respostes, les oracions declaratives, i les expressions de satisfacció o de descontentament. Finalment, el quart

estudi investiga la reacció dels infants a diversos tipus de combinacions de parla i gest d'assenyalar, i mostra que els infants de 12 mesos utilitzen les marques prosòdiques i gestuals del discurs per a entendre les intencions comunicatives que els són dirigides. En conjunt, aquesta tesi mostra que la integració temporal de prosòdia i gest ocorre en les etapes més primerenques del desenvolupament lingüístic i cognitiu, i que el usos pragmàtics de la prosòdia i la gestualitat emergeixen abans que els infants dominin l'ús d'elements lèxics. Així, la prosòdia és el primer component gramatical del llenguatge que els infants utilitzen amb finalitat comunicativa, cosa que indica que la comunicació lingüística emergeix abans que els infants tinguin la capacitat de produir ítems lèxics amb significants semàntics. La conclusió general és, doncs, que la integració temporal i pragmàtica de la prosòdia i el gest per part dels infants indica el desenvolupament primerenc de la pragmàtica lingüística.

List of original publications

CHAPTER 2

Esteve-Gibert, N. & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication,* 57, pp. 301–316.

CHAPTER 3

Esteve-Gibert, N., Prieto, P., & Pons, F (submitted). Nine-month-old infants are sensitive to the prosodic and gesture alignment. *Infant Behavior and Development.*

CHAPTER 4

Esteve-Gibert, N. & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *Journal of Child Language, 40*(5), pp. 919–944.

CHAPTER 5

Esteve-Gibert, N., Prieto, P., & Liszkowski, U. (submitted). Prosody and gesture help infants to interpret social intentions. *Developmental Science*.

Table of contents

1. INTRODUCTION

1.1. The emergence of intentional communication

Immediately after birth infants engage in social behaviors. Infants imitate adults' facial expressions only some hours after being born (e.g. Meltzoff & Moore, 1983; Meltzoff & Moore, 1997), have a preference to look at human faces (e.g. Cassia, Turati, & Simion, 2004; Farroni et al., 2005; Johnson, Dziurawiec, Ellis, & Morton, 1991; Simion, Macchi Cassia, Turati, & Valenza, 2001), and engage in proto-conversations with turn-taking (e.g. Levinson, 2006; Murray & Trevarthen, 1986; Striano, Henning, & Stahl, 2006). It has been suggested that infants are born with a capacity to detect ostensive communicative signals directed to them and that they are innately adapted to decode them (Csibra, 2010; Csibra & Gergely, 2009).

However, in these early social behaviors the infants do not see their interlocutor as an intentional agent. Researchers agree that it is not until infants reach 9-12 months of age that their communicative acts become truly intentional and that they see the actions of their interlocutors as intentional (e.g., Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Piaget, 1953).

According to Tomasello (1995), two abilities indicate the start of intentional communication in infants: (1) the ability to distinguish means and goals in the infants' own action productions and in the

others' action productions, and (2) the infants' ability to engage in joint attention frames.

The ability of an infant to distinguish between means and goals in their own productions and in those of their interlocutors implies that the infant can differentiate between an action (the means) and the intention motivating this action (the goal). Literature on action production and action perception has found that at around 6-9 months of age infants start to understand actions as goal-directed and are able to distinguish intentional from accidental actions (e.g., Carpenter, Akhtar, & Tomasello, 1998; Jovanovic & Schwarzer, 2007; Meltzoff, 1995; Woodward, 1998, 1999) (see Hauf, 2007, for a complete review). Woodward (1999) tested 5- and 9-month-old infants in four conditions to see if at these ages infants could distinguish between goal-directed actions and non-goal-directed actions. In two conditions, infants saw an actor intentionally grasping one object (called 'grasp condition') or accidentally touching an object with the back of their hand (called 'back-of-hand condition'). Then the test trials differed from the habituation trials with respect of showing a new object (called 'new object condition') or a new path to touch the object (called 'new path condition'). The authors analyzed the infants' looking times and found that at 9 months of age infants reacted differently to the goal-directed grasping action than to the non-goal-directed action, and that they preferred the new object condition over the new path condition. Results with 5-month-old infants showed similar but weaker patterns, suggesting that at that age the ability to distinguish intentional from non-intentional actions is starting to emerge but is

still not fully developed. Infants' ability to distinguish intentional from non-intentional actions was also tested by Carpenter, Akhtar and Tomasello (1998) to see if infants could base this distinction on the basis of the vocal behavior accompanying the action. These authors showed that 16-month-old infants imitated more intentional than accidental actions, the only difference between the two being that intentional actions were accompanied by the word 'There!' while accidental actions were accompanied by the word 'Whoops!' (both words with the corresponding prosodic features).

The emergence of joint attention skills is another crucial step in the development of the child as a social agent (e.g., Bruner, 1975; Carpenter & Liebal, 2011; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Mundy & Newell, 2007; Mundy et al., 2007; Repacholi & Meltzoff, 2007; Senju & Csibra, 2008; Tomasello & Farrar, 1986; Tomasello, 1995). One of the first experimental studies investigating the emergence of joint engagement is Trevarthen and Hubley (1978), in which the authors analyzed the behavior of one child who was interacting with a caregiver in a laboratory setting during the infant's first year of life. The authors found that before 6 months of age the infant interacted with the object alone or with the adult alone, but never transferred smiles or gazes from one to the other. The infant's behavior at this stage was thus characterized by dyadic interactions in which the infant did not see the other person as an intentional actor. However, from 6 to 9 months of age the infant showed the first sharing behaviors by smiling or looking at the mother while manipulating the object. And from 10 months onwards the infant frequently

alternated her gaze between object and adult and smiled at the adult in connection with the object, thus engaging in triadic interactions.

The way infants develop intention understanding skills has been widely studied in the literature on "theory of mind" abilities. Theory of mind refers to the ability to make predictions about another person's behavior and to infer one's own and others' mental states, i.e., the others' intentions, beliefs and desires (Premack & Woodruff, 1978, and many others thereafter). Infants at around 9 months of age distinguish between goal-directed and non-goal-directed actions (Woodward, 1998, 1999). But infants also have to learn to infer the other's belief and desire motivating the intention of the action in order to have "mind-reading" abilities. Theory of mind abilities have been typically assessed by means of false-belief tasks, such as the Sally-Ann task (Bahron-Cohen, Leslie, & Firth, 1986; Wimmer & Perner, 1983), in which the child observes a scene where a first agent places an object in a container and then leaves. While the first agent is absent, a second agent moves the object to a second container. When the first agent comes back, the child is asked where he/she will look for the object. The child will pass the task if (s)he points at the container where the agent left it before leaving, but will fail the task if (s)he points at the container where the second agent placed the object. Experimental studies using this false-belief task have found that children look at the right container at 3 years of age and that they explicitly point to the right one at 4 or 5 years of age (e.g., Clements & Perner, 1994; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

Passing false-belief tasks requires complex mental abilities such as inhibition processes (inhibiting one's own perspective to generate a different one while holding the relevant perspective in working memory) (Carlson, Claxton, & Moses, 2014; Russell, 1997), shared neural representations (directly accessing their own and the others' psychological states, which must be reflected in neural systems) (Gallese & Goldman, 1998; Goldman, 2009), or theorizing about the relation between mental states and actions (forming abstract concepts about such mental states and actions) (Gopnik & Wellman, 2012; Gopnik, 2003; see Mahy, Moses, & Pfeifer, 2014 for a complete review of the three accounts). Due to this complexity, it has been claimed that false-belief tasks should not be the only measure of theory of mind abilities (Carlson et al., 2014; Moses & Tahiroglu, 2010). Besides, mind-reading abilities also imply the ability to cognitively infer others' emotions. Emotion-detection tasks have been found to be correlated with intention-understanding, and both abilities are processed in the same regions of the brain (Brüne, 2005; Buitelaar & van der Wees, 1997; Henry, Phillips, Crawford, Ietswaart, & Summers, 2006; Mier et al., 2010).

Several studies have created less cognitively demanding tasks that investigate whether younger children show evidence of mind-reading abilities (Baillargeon, Scott, & He, 2010; Buttelmann, Carpenter, & Tomasello, 2009; Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005). As a whole, these studies show that infants younger than 3 or 4 years of age show mind-reading abilities when tested with cognitively easier tasks. Onishi and Baillargeon (2005) tested 18-month-olds in a violation-of-expectation

experiment. In this experiment infants were familiarized with a task in which an agent hid a toy inside a box and then this toy either moved to the other box in the agent's absence or moved to the other box while the agent was present but then returned to its initial position when the agent had left. In the test trial, the agent appeared either looking for the object in either one box or the other, and the infants' looking time to the event was measured. Results showed that infants looked longer if the agent tried to find the toy where the agent was not expected to do so on the basis of what (s)he had previously seen, revealing that at even such a young age infants make predictions of others' beliefs.

Within the studies that propose an early development of the theory of mind, two distinct but related approaches are found, nativist accounts and usage-based accounts. Nativist accounts of mind-reading abilities propose that humans are born with an innate ability to understand others as social agents with intentions (e.g., Kovács et al., 2010; Onishi & Baillargeon, 2005). Kovács et al. (2010), for instance, tested 7-month-olds in an object detection task to investigate whether at such a young age infants could make automatic computations of an agent's attention to objects. In a series of studies, the authors found that the 7-month-old infants' online reactions were influenced by an agent's attention to objects even if the presence of this agent was irrelevant to the action. They showed that the infants' looking times were longer when their own expectation of a future event was not confirmed and, crucially, also when an agent's expectation of a future event was not confirmed.

These results provide evidence of some sort of automatic belief computation and a human-specific 'social sense'.

Usage-based accounts suggest that very young infants learn that others are intentional agents through the social interaction experience, which triggers the system that predicts actions. In this regard, several studies provide evidence that infants have flexible expectations about other people's behaviors, and that these expectations depend on the social contextual situation in which such behaviors occur. Specifically, they show that young infants react flexibly depending on the others' behavior (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Liszkowski, Carpenter, & Tomasello, 2008; Moll & Tomasello, 2007; Liszkowski, 2013; Southgate, Chevallier, & Csibra, 2010), that they also initiate actions flexibly depending on the others' behavior (Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004; Liszkowski, Carpenter, & Tomasello, 2007; Liszkowski, Schäfer, Carpenter, & Tomasello, 2009), and that they intervene proactively to modify the others' behavior (Knudsen & Liszkowski, 2012a, 2012b).

In sum, research on the emergence of intentional communication shows that from the second half of the first year infants are able to transmit their intentions to others, and that at that moment they also start inferring the others' intentions, beliefs, and desires. These studies found that the social contextual information preceding an intentional act is crucial for infants to show that their behavior is intentional and to understand the intentional value of the others' acts. However, less research has been done on the linguistic and

gestural cues that accompany the actions and young infants might use them for intentional communication, and the present dissertation will shed some light on this issue.

1.2. Prosody in early intentional communication

Infants' sensitivity to prosodic features starts very early in language development, as evidenced by perception studies. Infants prefer the prosodic properties of infant-directed speech to those of adult-directed speech (e.g., Fernald & Kuhl, 1987; Fernald, 1985). Fernald (1985) found that 4-month-old infants preferred to hear infant-directed speech versus adult-directed speech, the difference between the two being their prosodic features, and Fernald and Kuhl (1987) investigated whether this preference was driven by the characteristic fundamental frequency (F0) patterns (higher pitch range values), duration patterns (longer values), or amplitude patterns (less variability). The authors found that infants' preference for infant-directed stimuli was motivated by certain prosodic cues: infants preferred the higher pitch range values of infant-directed speech, but they did not show a preference for its duration or amplitude features.

The early sensitivity to the prosodic patterns is also evidenced by the fact that 3-month-old infants have the ability to distinguish two languages if they belong to distinct rhythmic categories (e.g., Mehler, Jusczyk, & Lambertz, 1988; Nazzi, Bertoncini, & Mehler, 1998), a property that is shared with other animal species (e.g.,

Ramus, Hauser, Miller, Morris, & Mehler, 2000; Toro, Trobalon, & Sebastián-Gallés, 2003). Also, 4- to 5-month-old infants can distinguish languages within the same rhythmic category if these languages have distinct segmental cues (Bosch & Sebastian-Galles, 2001; Nazzi, Jusczyk, & Johnson, 2000). Finally, infants are also sensitive to the position of prosodic prominence very early on and at 6-9 months of age they prefer the stress pattern of their home-environment language (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk, Cutler, & Redanz, 1993; Pons & Bosch, 2010).

In terms of early production of prosodic patterns, infants are found to develop distinct dimensions of prosody (i.e., rhythm, stress, and intonation) at distinct stages. Mampe, Friederici, Christophe, and Wermke (2009) investigated the intonation and intensity contours of the first cries of French and German newborns (mean age of 3 days) to see if they differed as a function of the language they heard prenatally. Their results showed that all infants cried following a rising-falling arch-shape, as expected, while the peak of the arches in the melody and intensity contours varied significantly depending on the language they had been exposed to: French infants' cries showed the peak of the melody and intensity contour towards the end of the arch, thus displaying a rising contour, while German infants' cries produced the peak of the contours at the beginning of the arch, thus displaying a falling contour.

Stress patterning is found to develop a bit later in the process of language acquisition (e.g., Behrens & Gut, 2005; Davis,

MacNeilage, Matyear, & Powell, 2000; DePaolis, Vihman, & Kunnari, 2008; Keren-Portnoy, Majorano, & Vihman, 2009; Snow, 2006; Vihman, DePaolis, & Davis, 1998; Vihman, Nakai, & DePaolis, 2006). Davis et al. (2000), for instance, analyzed the acoustic parameters of stress in babbling infants and found that, although they were able to use fundamental frequency, intensity, and duration to signal prominence, at that period they did not produce the acoustic cues of stress in an adult-like way. Vihman et al. (1998), however, investigated French- and English-learning infants and found that at the 25-word point (i.e., the one-word stage) French infants were producing more iambic words than trochees (like adults do), while English infants produced both types of patterns. The authors suggested that at that stage infants are already tuned to their ambient language, since in French all words are stress-final while in English there is a preference for stress-initial words –although both patterns are found. However, it has been found that it is not until children are much older that they acquire the rhythmic patterns of the ambient language. Payne, Post, Astruc, Prieto, and Vanrell (2011) compared the rhythmic patterns of 2-, 3-, and 4-year-old children in Catalan, Spanish, and English. The authors found that there is some evidence that at age 2 children use some rhythmic cues consistently with the ambient language (particularly interval variability), although results improved significantly across the subsequent ages.

But prosody is not merely an acoustic feature. One of the primary roles of prosody is to contribute to the pragmatic meaning of the utterance. Prosody is used to express the speaker's attitude towards

an object or event, to distinguish among sentence types, to structure information, to organize and maintain interactions, or to convey epistemic and evidential information (see Barth-Weingarten, Dehé, & Wichmann, 2009, for a review of the prosody-pragmatics interface). Several studies on the development of the prosodic-pragmatics interface report that the emergence of a complex inventory of intonation contours with an adult-like intentional meaning occurs around the two-word stage (Chen & Fikkert, 2007; Frota & Vigário, 2008; Prieto, Estrella, Thorson, & Vanrell, 2012). And other studies on children's comprehension of the pragmatic meaning of prosody have revealed that the comprehension of prosodic features at sentence level is acquired quite late in language development (e.g., Cruttenden, 1985; Cutler & Swinney, 1987; MacWhinney, Pléh, & Bates, 1985). In fact, Cutler and Swinney (1987) claim that there is a 'performance paradox' because some prosodic contours are produced that are appropriate for the intentional meaning but at the same age infants have trouble understanding the meanings these contours convey in comprehension tasks.

However, it is reasonable to think that infants use their early sensitivity to the acoustic cues of prosody well before the two-word stage to comprehend intentional meanings and communicate with others. Two observations motivate this assumption. First, the fact that infants comprehend and produce intentional actions already during the second half of the first year (Woodward, 1998, 1999). Second, the fact that infants are able to use prosodic cues consistently some months after birth, not only in terms of

perception of the acoustic features (Bosch & Sebastian-Galles, 2001; Fernald & Kuhl, 1987; Jusczyk et al., 1993; Mehler et al., 1988; Nazzi et al., 1998; Ramus et al., 2000) but also in terms of production of non-meaningful speech (Mampe et al., 2009).

In fact, some studies suggest that young infants might be able to use prosody for early intentional communication (D'Odorico & Franco, 1991; Papaeliou & Trevarthen, 2006; Papaeliou, Minadakis, & Cavouras, 2002; Sakkalou & Gattis, 2012). In a comprehension study, Sakkalou and Gattis (2012) tested whether 14- and 18-month-old infants would imitate more intentional than accidental actions, when the difference between the two actions was the prosody of the word accompanying the actions (with methods based on Carpenter et al., 1998). In a first experiment intentional actions were accompanied by the word 'There' with high amplitude and long duration, while accidental actions were accompanied by the word 'Whoops!' with low amplitude and short duration; in a second experiment they used the same methodology but removed the lexical information. In both experiments the authors found that infants imitated the intentional actions more often than the accidental ones, and they observed an age-related effect when lexical cues were removed, namely older infants performed better than younger infants. These results suggest that infants comprehend the pragmatic value of prosodic cues and are able to relate them to intentionally at 14 months of age since they could only base their imitative behavior on the prosody accompanying the action.

Papaeliou and Trevarthen (2006) investigated the early production of prosodic cues to intentionality. Specifically, they aimed to see whether acoustic parameters (the duration of the vocalization, the beginning, final, maximum, minimum, and mean fundamental frequency values, and the range and standard deviation of fundamental frequency) reflected the intentional or non-intentional value of vocalizations. They recorded four 10-month-old English-learning infants in two distinct situations: while playing with the mother and while playing alone. Vocalizations were considered to be intentional when produced while directing gaze to the mother, producing communicative gestures, or following the mother's gaze and pointing gestures, or as a consequence of behaving as the mother required; vocalizations were considered to be non-intentional when produced while holding an object, inspecting an object, or completing a preceding behavior. Results confirmed their hypothesis and showed that intentional vocalizations were shorter and with higher pitch values than non-intentional vocalizations.

Despite these two studies exploring the role of prosody in the early production and comprehension of pragmatic meanings, much more research needs to be done to clearly understand the emergence of the link between prosody and pragmatics. These previous studies have found that infants distinguish intentional from non-intentional meanings by means of the production and comprehension of different prosodic patterns. However, it is not yet known whether prosody is used by young infants to convey and interpret specific pragmatic meanings in their social interactions.

1.3. Gestures in early intentional communication

One clear behavioral manifestation of the infant becoming an intentional agent is the production and understanding of pointing gestures. The pointing gesture is a simultaneous extension of the arm and index finger towards a target with the aim of directing the attention and behavior of another person to an object or event. Together with reaching gestures (extension of the arm and opening of the hand towards an entity in order to direct the caregiver's attention to it), they form the category of deictic gestures. Pointing gestures are considered a clear sign of intentional communication because their production implies that the speaker has the goal of redirecting the other's attention, and its understanding implies that the interlocutor recognizes the other as having the goal of directing their attention (Bates, Camaioni, & Volterra, 1975; Kita, 2003; McNeill, 1992).

Distinct social intentions[1] have been identified as underlying the act of pointing towards an object or an event. Bates et al. (1975) distinguished between declarative pointing and imperative pointing: declarative pointing was used to get the adult to attend to an external entity using the external entity as a tool to obtain the

[1] Throughout this dissertation, the terms 'social intention', 'pragmatic meaning', and 'communicative intention' are used interchangeably to refer to the meaning conveyed in the communicative act that is produced in a context of social interaction and is intended for an interlocutor.

adult's attention, while imperative pointing was used to get the adult to retrieve an object for them, using the adult as a tool to obtain the object. Building on Bates et al. (1975), Tomasello, Carpenter, and Liszkowski (2007) proposed that pointing was a communicative act that directs the interlocutor's attention towards an object with three main social intentions: (1) helping the interlocutor with some information that might be of interest to them (*declarative informative pointing*), (2) sharing attention with the interlocutor about an object or event (*declarative expressive pointing*), and (3) requesting an object from the interlocutor (*imperative pointing*).

Imagine for example that Ann and James are talking to each other, Ann facing a window and James with his back to it. During the conversation it starts raining and Ann points towards the window to shift James' attention towards it in order to inform him that something relevant for him is happening out there. That would be an example of declarative informative pointing. Now imagine that you and I are watching fireworks and looking open-mouthed at the show. I point towards the sky while you are looking at it because I want to share my attention with you with respect to a particular pattern that a firework has created. That would be an example of declarative expressive pointing. And now think of a situation in which you are having lunch together with your family; your mother points towards the bottle of wine because she wants you to pass it but does not speak because she is eating something. That would be an example of imperative pointing.

The comprehension of the meaning of pointing gestures has been investigated in several studies, revealing that infants can successfully interpret the intention of a pointing gesture when the social action context gives them sufficient information to do so (Aureli, Perucchini, & Genco, 2009; Behne, Carpenter, & Tomasello, 2005; Behne, Liszkowski, Carpenter, & Tomasello, 2012; Camaioni, Perucchini, Bellagamba, & Colonnesi, 2004). Camaioni et al. (2004) showed that infants modified their reactions when a pointing gesture was directed at them with either an imperative or a declarative expressive intention, thus demonstrating that they understood the difference between the two intentions motivating the gesture. However, they found age-related differences: imperative points were understood at 12 months of age and expressive points at 15 months of age. Behne et al. (2012), however, found that already at 12 months of age infants could interpret correctly the informative nature of a pointing gesture directed at the hidden location of a toy, and that comprehension scores were correlated with production scores.

Infants start producing communicative pointing gestures around 10-12 months of age, both declaratively and imperatively. Cochet and Vauclair (2010) studied extensively many aspects of pointing development in 15- to 30-month-old infants with three tasks intended to elicit declarative informative, declarative expressive, and imperative pointing gestures. Their results showed that: (a) declarative expressive and informative points were more frequently accompanied by vocalizations than imperative points, supporting other findings that declarative points might be more connected with

language development than imperative points (Camaioni, Perucchini, Muratori, Parrini, & Cesari, 2003; Camaioni et al., 2004); (b) declarative points lasted longer than imperative points, suggesting that the infant was trying to maintain interactions in the declarative situation (although the authors noted that these results might be affected by other variables in the interaction); and (c) hand shape distinguished between declarative and imperative points, with declarative points involving the index finger and imperative ones involving the whole hand.

In a series of studies, Liszkowski et al. (2004, 2006) investigated whether infants at 12 months of age were able to point with an informative and expressive social intention given the appropriate social context. In Liszkowski et al. (2004), an experiment was designed in which an adult reacted differently to the infants' points (i.e., with joint engagement, by only looking at the infants' face, by only looking at the event, or by ignoring both the infant and the event). Their results showed that 12-month-old infants pointed to share attention with the adult with a declarative expressive intention because they pointed more frequently in the joint engagement condition than in the other conditions. Liszkowski et al. (2006) tested 12- and 18-month-old infants in their ability to use a pointing gesture to inform an adult about the location of a dropped object. Confirming their expectations, they found that at both ages infants were able to point informatively.

Importantly, the early production of pointing gestures has been found to correlate positively with later linguistic and grammatical

abilities. Igualada, Bosch, and Prieto (2014) confirmed and extended previous findings by Murillo and Belinchón (2012), revealing that infants who produced more pointing gestures in combination with speech at 12 months of age were the ones with more vocabulary and better grammatical development at 18 months of age. Iverson and Goldin-Meadow (2005) and Ozçalişkan and Goldin-Meadow (2005) analyzed infants during their transition between the one- and two-word stages, and found that one particular type of gesture-speech combination, namely that in which the gesture provides supplementary meaning to speech, predicted grammatical development. And Rowe and Goldin-Meadow (2009) found that the number of gesture-speech combinations at 18 months of age predicted sentence complexity at 42 months of age.

1.4. Prosody and gesture integration in adults

As adult speakers, most of our communicative gestures are produced not in isolation but combined with speech. Speech and gesture work together from a phonological and pragmatic point of view to convey the speaker's intended meaning in human communication (e.g., Kendon, 1980; McNeill, 1992). The phonological alignment between gesture and speech refers to the fact that the most prominent part of the gesture (whether it is the 'gesture stroke' or the 'gesture apex', see Figure 1)[2] coincides with

[2]According to McNeill (2005), the stroke phase of a gesture is typically the interval of apparent greatest gestural effort, and 'effort' is determined with

the most prominent part of speech, generally identified as the stressed or accented syllable in the utterance accompanying the gesture (e.g., De Ruiter, 2000; Levelt, Richardson, & La Heij, 1985; Loehr, 2012; Nobe, 1996; Rochet-Capellan, Laboissière, Galván, & Schwartz, 2008; Rusiewicz, 2010; Yasinnik, Renwick, & Shattuck-Hufnagel, 2004). Loehr (2012) analyzed natural interactions of English speakers to see if gesture and prosody were aligned at several levels. Following McNeill (1992), the levels taken into account for the gesture analysis were gesture apices, gesture phases, gesture phrases, and gesture units. Following the autosegmental-metrical system (Pierrehumbert, 1980), the levels taken into account for the prosodic analysis were pitch accents, intermediate phrases, and intonational phrases. The author found that gesture apices reliably predicted the presence of a pitch accent, and that gesture phrases reliably correlated with intermediate phrases.

Figure 1. Phases of a pointing gesture: (1) preparation phase; (2-4) gesture stroke; (3) gesture apex; (5) retraction phase.

---

reference to parameters such as relative forcefulness of movement or apparent tenseness of hand shapes. In the case of pointing gestures, for instance, the gesture stroke is an interval of time in which the arm is maximally extended, and the gesture apex is the specific point within the stroke interval at which the finger is maximally extended.

Prosody and gesture structures are also found to mutually influence each other in perception and production tasks. Krahmer and Swerts (2007) investigated the influence of the presence of a visual beat gesture[3] on the perception and production of prosodic prominence. The authors found that producing a visual beat alters the acoustic realization of the prosodic prominence, and that seeing a visual beat makes speakers perceive greater prosodic prominence. The influence of the prosodic structure on gestural timing was investigated in Esteve-Gibert and Prieto (2013) in a task in which participants had to point while producing a target word in a contrastive focus position. The authors found that the position of the gesture apex depended on the position of the intonation peak within the accented syllable, which at the same time was influenced by the presence of a preceding or upcoming phrase boundary.

The pragmatic (and semantic) alignment of gesture and speech refers to the fact that the meaning conveyed in gesture parallels or contributes to the meaning conveyed in speech. Kelly, Ozyurek, and Maris (2010) tested speakers' comprehension of concepts in an experimental setting. Participants saw various actions represented by speech and gesture modalities in three different conditions: when both modalities conveyed congruent information (speech: "chop"; gesture: chop), when both modalities conveyed slightly incongruent

[3]A beat gesture is a biphasic movement with the hands or head parts that does not present a discernible meaning but merely accompanies speech.

information (speech: “chop”; gesture: cut), and when both modalities conveyed totally incongruent information (speech: “chop”; gesture: twist). Their results showed that the understanding of the actions was affected by the level of incongruency (the higher the incongruency, the greater the difficulty in understanding), and that the influence of gesture on speech was omnipresent. Several neuroimaging studies have also confirmed the semantic and pragmatic integration between gesture and speech, revealing that the brain responds differently depending on whether gestures match or mismatch the semantic content of the lexical items (Habets, Kita, Shao, Ozyurek, & Hagoort, 2011; Ozyurek, Willems, Kita, & Hagoort, 2007; Willems, Ozyurek, & Hagoort, 2009).

One particular aspect of speech, prosody, has been revealed as crucially complementing or supplementing the meaning of gesture cues. Crespo-Sendra, Kaland, Swerts, and Prieto (2013) investigated the use of prosodic and facial gesture features by Catalan and Dutch speakers while producing two sentence types, namely information-seeking questions and counter-expectation questions. The authors found that Catalan speakers used more facial gestures to mark sentence type than Dutch speakers, and they related this finding with the fact that Dutch uses more intonational strategies to distinguish sentence types. Thus, when a particular language does not have clear prosodic cues to distinguish between sentence types, gesture strategies enhance the distinction. Similar findings were obtained by Borràs-Comes, Kaland, Prieto and Swerts (2013) when comparing information-seeking questions and broad focus statements in Dutch and Catalan. In both languages speakers

benefited from facial gestures accompanying the speech, but Dutch speakers relied more on the audio cues than Catalan speakers, due to the fact that Dutch uses syntactic strategies to distinguish between sentence types.

The tight temporal, semantic, and pragmatic alignment between gesture and speech has led various authors to propose several gesture production models that try to explain this fact. McNeill's (1992) Growth Point Theory regards gestures as communicative devices and claims that gesture and speech originate from the same mental imagery and are actually part of the same single system in communication. The Sketch Model (De Ruiter, 2000) differs from McNeill's account in that De Ruiter does not agree that gesture and speech come from the same imagery. Rather, De Ruiter proposes that gesture and speech are generated by separate systems but that they interact at an early stage of the speech production where the communicative intent is planned. The author claims that the realization of the gesture is planned before speech, and thus it is the gesture that influences speech and not the other way around. But not all accounts see gesture as communicative devices. Krauss, Chen and Chawla's (1996) model claims that gestures are not communicative and that their function is merely to facilitate lexical access. These authors also propose that the temporal integration of gestures and prosody is a result of the articulation stage, where the phonological encoder influences the motor movements. Finally, Kita and Özüyrek (2003) propose what they call the Interface Model, in which they claim that the two modalities come from different systems, that these systems interact during the different

stages of message formulation, and that the process of language encoding impacts on the formulation of the gesture.

1.5. Prosody and gesture integration in infants

Little is known about the development of the phonological and pragmatic coordination between speech and gesture in infants. Iverson and Thelen (1999) proposed four phases to explain the dynamics of the entrainment between gesture and speech in infants. These authors claim that in the first months of life infants show *initial linkages* between oral and manual systems. This is evidenced by the fact that infants bring their hands to the facial area and introduce their fingers into their mouth, open their mouth when a pressure is applied to their palms (the so-called Babkin reflex), and at around 2 months of age they start bringing objects to their mouth. After that, when they are 3 or 4 months of age and especially from 6 months onwards, they show *emerging control* between the two systems. Infants start producing rhythmical movements of the hands and arms that are timed together with vocal cooing and babbling. Later, around 10 or 11 months of age, infants show *flexible coupling* of gesture and speech, using gesture and speech for a communicative purpose. The authors state that this phase is characterized by an asymmetry between control and usage of the two modalities, with infants using gestures more frequently than speech to convey intended meaning, and by the uses of gesture predicting later language and grammatical development. Finally,

synchronization coupling emerges, with infants coordinating both modalities for intentional communication in an adult-like way, i.e., coordinating the most prominent part in gesture with the most prominent part in speech.

The integration of audio-visual (A-V) cues is present very early in infants' cognitive development. Several studies involving speech-accompanied articulatory gestures show that A-V temporal integration occurs very early in development. Very young infants can detect an A-V asynchrony of 500 ms when visual information (i.e., articulatory gestures) of an audiovisual event precedes the auditory (speech) information (Lewkowicz, 2010; Pons & Lewkowicz, 2014). Lewkowicz (2010) showed that 4- to 10-month-old infants detect an A-V asynchrony in an articulated syllable when the timing lag is 366 ms (speech preceding gesture), but only if they have been previously exposed to greater asynchronies. A further follow-up study by Pons and Lewkowicz (2014) showed that 8-month-old Catalan and Spanish infants are sensitive to the asynchrony of A-V events in fluent speech when the audio stream precedes the video stream by 366, 500, or 666 ms and that this effect is independent of their prior language experience. This early perceptual ability has been found to be relevant for infants to discriminate between articulatory gestures and phonetic contrasts, and to segment speech (Hollich, Newman & Jusczyk, 2005; Teinonen, Aslin, Alku & Csibra, 2008; Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastián-Gallés, & Werker, 2007).

The emergence of the temporal integration of gesture and speech has been also explored from a production point of view. Butcher and Goldin-Meadow (2000) analyzed the gesture and speech productions of six English-learning infants while they interacted naturally with an adult. They found that at the end of the one-word period infants combined gestures with speech, but that at that point they still did not temporally align the prominence in gesture with the prominence in speech; instead, this tight temporal alignment was only found at the two-word stage. Despite these interesting results, only inconclusive claims could be drawn from this data because the sample was not homogeneous and the tight temporal alignment was not analyzed following recent results regarding multimodal prominence alignment. Importantly, other studies on the emergence of pointing-speech combinations have reported the importance of these multimodal productions in the infants' later linguistic and grammatical development (Igualada et al., 2014; Iverson & Goldin-Meadow, 2005; Murillo & Belinchón, 2012; Özçalişkan & Goldin-Meadow, 2005; Rowe & Goldin-Meadow, 2009).

Despite the results presented above, previous studies do not offer a complete picture of how gesture and speech (and particularly prosody) develop in infants so that both modalities work together to convey intentional meaning. No research has explored infants' sensitivity to the alignment between gesture and speech with communicative gestures such as pointing gestures, in which the alignment is characterized by the coordination of the respective prominences. Similarly, no research has been done on whether

infants rely on the integration between gesture and prosody to produce and comprehend intentional communication at an early stage of linguistic and cognitive development. We know that (a) 10-month-old infants use phonetic cues of prosody to signal intentionality in their vocalizations (Papaeliou & Trevarthen, 2006), that (b) 14-month-old-infants rely on prosody to distinguish between intentional and non-intentional actions (Sakkalou & Gattis, 2012), and that (c) 14- to 18-month-old infants understand the purpose of pointing gestures if the context indicates the social intention (Aureli et al., 2009; Behne et al., 2012; Camaioni et al., 2004; Liebal & Tomasello, 2009). But as far as we know no previous studies have explored whether infants use gestures and prosody in an integrated way to convey specific social intentions and to understand these social intentions independently of the social contextual information.

1.6. General objectives, research questions and hypotheses

In this dissertation we aim at investigating how young infants integrate prosody and gesture for intentional communication before they can use lexical means for this purpose. Specifically, we are interested in whether young infants temporally integrate prosodic cues and pointing gestures from a perception and production point of view, and whether they successfully use this integration to

communicate intentionally and to comprehend intentional communication.

Four main research questions will be addressed, each one in a separate chapter:

1) Do young infants temporally align prosody and gestures in the context of intentional communication?
2) Are young infants sensitive to the temporal alignment between prosodic and pointing gesture prominence?
3) Do infants use prosodic and gesture means to signal intentionality and to express specific social intentions before using lexical cues?
4) Are young infants able to understand the others' intentions when these intentions are conveyed by means of prosodic and gesture shape in an integrated way?

Our hypotheses are that (a) infants are able to temporally integrate gesture and prosody from the beginning of pointing-speech combinations, and that (b) before producing pointing-speech combinations, they are sensitive to the fact that the two modalities have to be tightly aligned. We further hypothesize that (c) infants use this multimodal integration to convey specific social intentions, and that (d) they also use the integration of prosody and gesture for early intention understanding. The dissertation is thus organized in four independent studies, which are presented in Chapters 2 to 5. The first two studies (Chapters 2-3) investigate the early temporal integration between prosody and pointing gestures. The last two studies (Chapters 4-5) investigate infants' early use of the prosodic

and gesture integration for intention understanding and intention conveyance.

The first study (**Chapter 2**) focuses on how infants temporally align prosody and gesture when communicating intentionally. Previous studies have reported that at the transition between the one-word and two-word stages infants temporally integrated gesture and speech, but the temporal alignment was not precisely measured and the sample was not homogeneous. In order to correct these issues, we analyzed the gesture and acoustic cues of the intentional acts produced by four children from the babbling stage to the late one-word period (from 11 to 19 month old) while interacting normally with their parents. The results showed three main findings: (a) the infants combined gesture and speech from the onset of word production; (b) most of the combined gestures were pointing gestures with a declarative intention; and (c) these combinations showed an alignment pattern that is very close to the one observed in adults (in that gestures precede speech, the onsets of the gesture and prosodic prominences coincide, and the apex of the gesture occurs before the end of the prosodic prominence).

But in order to have a complete picture of how infants develop integration between gesture and prosody from a temporal point of view, we needed to investigate this issue from a perception point of view. In the second study (**Chapter 3**) we were therefore interested in early infants' sensitivity to the temporal alignment between gesture and prosodic prominences. Previous research had shown that 4- to 10-month-old infants are sensitive to the temporal

alignment between articulatory gestures and their corresponding speech, but no previous research has investigated whether children are sensitive to the integration of prosody and gesture. Using a head-turn paradigm, we tested 9-month-old infants to see whether they perceived the misalignment between gesture and prosodic prominence, i.e., whether they could detect it when the stroke of the gesture did not coincide with the accented syllable in co-speech pointing gestures. The results showed that already at 9 months of age the infants were sensitive to the temporal alignment of both the two modalities.

In the third study (**Chapter 4**) we investigated whether young infants at a pre-lexical stage use prosodic cues (and gesture cues) to communicate intentionally. Previous findings had shown that infants use prosodic means to distinguish between intentional and non-intentional speech acts, but there was a lack of evidence about whether at this age infants were also able to use prosody and gestures to convey more specific social intentions. By analyzing a longitudinal corpus of four Catalan-babbling infants recorded at home during spontaneous interactions, we investigated whether children use different prosodic patterns to distinguish intentional from non-intentional vocalizations and to express specific intentions. Vocalizations from 0;7 to 0;11 were coded acoustically (i.e., for pitch range and duration), gesturally, and pragmatically. The results showed that prosodic cues were different for intentional and non-intentional vocalizations, and that prosody was also used to signal the specific pragmatic meaning of intentional vocalizations. Specifically, requests and expressions of discontent displayed wider

pitch excursions and longer durations, and statements and responses displayed narrower pitch ranges and shorter duration values. These results thus constitute some of the first evidence of an early link between prosody and pragmatics.

In the fourth study (**Chapter 5**) we investigated whether this ability to use prosody and gestures in early intentional communication was also found in comprehension. Previous research revealed that infants use the social contextual information to understand the speaker's intentionality, but we wanted to explore whether prosodic and gesture cues could also be relevant in the infants' understanding of the others' intentions. We designed two experiments to see whether 12-month-old infants could distinguish among the expressive, imperative or informative meanings of an attention-directing act using the gesture and prosodic cues produced by the speaker, controlling for either the social contextual information preceding the action (Experiment 1) and lexical cues (Experiment 2). Our results showed that the infants indeed understood the speaker's specific intention because they reacted mostly appropriately in each condition, and that gesture shape (either whole-hand or index-finger pointing) and prosodic cues (duration and pitch range) were crucial in this early understanding of intentional communication. Our findings thus showed that the prosodic and pointing gesture features accompanying the attention-directing act were also processed by infants at such an early age in order to understand the other's intent. This study is the first of its kind to show that social contextual cues are important but not

indispensable, because infants can use prosody and gesture shape to understand intentional communication in the absence of such cues.

CHAPTER 2: INFANTS TEMPORALLY ALIGN GESTURE-SPEECH COMBINATIONS BEFORE THEIR FIRST WORDS

2.1. Introduction

There is a broad consensus in the literature on the tight relationship and mutual influence between gesture and speech. Many researchers have stated that gesture and speech form an integrated system in communication (e.g. De Ruiter, 2000; Kendon, 1980; Kita, 2000; McNeill, 1992). Important features that back up the speech-gesture integration analysis in adults are that most of the gestures are produced together with speech, and that the two modalities are (a) semantically and pragmatically coherent, and (b) temporally aligned, i.e., the most prominent part of the gesture is temporally integrated with speech (McNeill, 1992).

Studies on the temporal alignment of gesture and speech provide strong evidence for the claim that gesture and speech form an integrated system in adults. It has been shown that the most prominent part of the gesture typically co-occurs with the most prominent part of the speech (Kendon, 1980). But different anchoring regions in speech have been proposed to serve as coordination sites for gesture prominence locations: speech onset (Bergmann, Aksu, & Kopp, 2011; Butterworth & Beattie, 1978; Ferré, 2010; Levelt et al., 1985; Roustan & Dohen, 2010), prosodically prominent syllables (Krahmer & Swerts, 2007; Leonard & Cummins, 2010; Loehr, 2012), or prosodically

prominent syllables with intonation peaks (De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Nobe, 1996). Taking together these findings, there is general agreement in the literature that (a) prominences in gesture and speech are temporally aligned, (b) the prominence in gesture is represented by the gesture stroke (in the case of a pointing gesture, the interval of time during which the arm is extended) or the gesture apex (the specific point within the stroke interval at which the finger is maximally extended), and (c) the prominence in speech is represented by the prosodically prominent syllable. In the present study, these measures will be taken into account in order to investigate the development of the temporal alignment of gesture with speech in the early stages in language development.

But are gestures aligned with speech in young infants to form an integrated system the way that they are in adults? Iverson and Thelen (1999) and Masataka (2003) stated that speech and gesture combinations have their developmental origins in early hand-mouth linkages. Based on the dynamic systems theory of motor control, they proposed that systems activating mouth and arms can influence and entrain one another, and these entrainments are dynamic and flexible such that activation in one system can affect the other in the form of a looser or tighter temporal synchrony. However, a given behavior must be strong and stable (with low threshold and high activation) to pull in and entrain the activity of the complementary system. Iverson and Thelen (1999) proposed four developmental stages of dynamic progression, namely, (1) in newborns, an early oral-manual system in which instances of hand-mouth contact and co-occurrences of hand movements with vocalizations are frequent;

(2) from 6 to 8 months, rhythmical movements with the hands and mouth showing an increasing control over the manual and oral articulators, and possibly indicating the transition into the speech-gesture system; (3) from 9 to 14 months, a more articulated control over the two modalities, which are then more directed to communication, with the gesture modality predominating but with entrainment also occurring between the two, and showing a tight relation between early gesture production and a later language development, and (4) from 16 to 18 months, a tighter control over both modalities, leading to the emergence of synchronous gesture and speech combinations.

In order to investigate deeply the temporal overlap between the occurrence of vocalizations and rhythmic activities of the limbs in infants, Ejiri and Masataka (2001) investigated the vocal and motor behavior of 4 Japanese infants from 6 to 11 months of age. In a first study, they examined the temporal overlap between vocalizations and rhythmic activities during the babbling stage. The authors found that vocalizations very frequently co-occurred with rhythmic actions, and interestingly that these coordinated behaviors increased immediately before and during the month in which canonical babbling was initiated. In a second study, they compared vocalizations co-occurring with rhythmic actions to vocalizations not co-occurring with rhythmic actions, and they found that syllable and formant frequency durations were shorter in vocalizations co-occurring with rhythmic actions than in non-co-occurring ones. Similarly, Iverson and Fagan (2004) described early infants' production of vocal-rhythmic movement coordination by testing 47

infants between the ages of 6 and 9 months. Results showed that at 7 months of age vocal-motor coordination was a stable component of infants' behavioral repertoires, and that these early combinations were a developmental precursor to the gesture-speech system. The authors based this statement on three observations: (1) infants at all ages coordinated vocalizations with single-arm rhythmic movements more often than with both-arm movements; (2) at all ages the proportion of coordinated right-arm movements was higher than that of left-arm movements, paralleling adult-like behaviors; and (3) most of the combinations followed the temporal patterns of organizing gesture-speech productions, since motor activities were synchronous with or slightly anticipated vocalization onsets.

The abovementioned studies focusing on rhythmic movements revealed that vocal and motor rhythmic movements are precursors of the alignment between the gesture and speech modalities. However, very few studies have investigated the patterns of that early alignment itself, i.e., the early alignment between communicative gestures and vocalizations. To our knowledge, only Butcher and Goldin-Meadow (2000) have performed such a study. The authors analyzed six infants longitudinally in spontaneous play situations during the transition from one- to two-word speech in order to investigate whether (1) at that age infants produce gestures with or without speech, (2) infants temporally align gesture and speech, and (3) infants semantically integrate the two modalities. First, they found that the production of utterances containing gesture remained stable across the ages analyzed, but with a difference between age groups: at the beginning of the single-word

period gestures were generally not accompanied by speech, and at the end of the single-word period infants mainly combined them with speech. Second, they found that it was not until the beginning of the two-word period that infants produced gesture-speech combinations in which the speech co-occurred with the most prominent part of the gesture (defined by them as the stroke or peak of the gesture, i.e., the farthest extension before the hand began to retract). Finally, the study showed that the proportion of gestures produced in combination with meaningful speech (as opposed to meaningless speech) increased across the ages analyzed. In conclusion, Butcher and Goldin-Meadow (2000) suggested that it is not until the beginning of the two-word period that infants integrate gesture and speech as a single system to communicate intentionally.

The present chapter aims at describing the emergence of gesture-speech combinations and their temporal alignment. Following up on Butcher and Goldin-Meadow (2000), we incorporate two innovative aspects in our study. The first innovative aspect is an analysis of the emergence of communicative gesture-speech combinations starting from the babbling period. The babbling period emerges in the middle of the first year of life and it is a crucial stage in language development because it provides the raw material for the production of early words (Oller, Wieman, Doyle, & Ross, 1976; Vihman et al., 1985). In the frame of the dynamic systems theory (Iverson & Thelen, 1999), Vihman, DePaolis, and Keren-Portnoy (2009) propose an ‘articulatory filter’: the first syllables that infants produce when babbling help the bootstrapping of the speech stream, and consequently the development of the phonological

systematicity. Thus, during the second half of the first year, infants start practicing very simple, accessible, and accurate syllables. Once these syllables are well practiced, the infants' attention is unconsciously captured by sound patterns in the speech stream that match good enough their own babbled productions. Consequently, the infant can detect if a sound pattern occurs repeatedly in a given situation and, when experiencing a similar situation, the infant will be primed to produce those syllables. According to the authors, this fact can strengthen the memory trace and support the memory for the mapping between form and meaning. Also, the babbling period coincides with the period when communicative gestures start being produced (Bates et al., 1975; Tomasello et al., 2007). The second innovative aspect in our study is a more fine-grained temporal alignment analysis that incorporates recent findings on the way gesture and speech temporally align in adult speech and that takes into account the importance of prosodic prominence in gesture-speech alignment patterns. This will allow us to assess the degree of temporal align in more detail.

Thus, the goal of this study is twofold. First, it aims to describe when and how infants combine communicative gestures with speech in the babbling and single-word periods. In order to fulfill this aim, the study will analyze the intentional gesture-speech combinations produced by 4 infants between 11- and 19-months of age, the ages in which infants start producing most of their communicative gestures in combination with speech, and then go onto examine the gesture types and motives that appear most frequently in these early gesture-speech combinations. Second, it

aims to investigate precisely the early temporal alignment of gesture with speech. To this end, the study will analyze a variety of measures that have been found useful in recent studies involving adults, as follows: the temporal distance between gesture onset and speech onset; the temporal distance between stroke onset and speech onset; the temporal distance between stroke onset and the beginning of the accented syllable; and the temporal distance between the gesture apex and the end of the accented syllable. We hypothesize that we will find evidence of temporal alignment in early gesture-speech combinations as they emerge in the transition between the babbling and single-word periods.

2.2. Method

2.2.1. Participants

The participants of the longitudinal study are four Catalan-learning infants, two male (who will be called Bi and Ma) and two female (who will be called An and On). The infants are all from middle-class homes in four small towns located within the same region of Catalonia, Alt Penedès, located 50 km to the south of Barcelona. Although varying degrees of bilingualism between Catalan and Spanish exist throughout Catalonia, according to the official statistics website of Catalonia (www.idescat.cat, Linguistic census from 2011) linguistic census, in that region Catalan is spoken

regularly by about 83% of the population. All parents of the four participants spoke exclusively in Catalan with their infant and to each other. Parents were asked about their linguistic habits through a language questionnaire, and they showed a mean 85% of use of Catalan in their dealings with other family members, friends, and work colleagues.

2.2.2. Procedure

The infants participated in free play activities as they interacted with their caregiver. Caregivers were told to interact naturally with the infants, playing as they would in their everyday lives. No other instructions on how to play or interact were given to them. Sessions were videotaped in the subjects' respective homes, typically in the living-room. The experimenter hand-held the camera while recording infant and caregiver, and if the infant -caregiver dyad moved to another room, the experimenter followed them with the camera. Recording sessions took place from when infants were 11-month-old until they were 19-month-old, either weekly or biweekly, and lasted between 30 and 45 min, depending on the attention span of the infants. All recordings have been made public through the Esteve-Prieto Catalan acquisition corpus, which includes recordings of these four infants from the age of 7 months until they were 3 years of age (Esteve-Gibert & Prieto, 2012). Recordings were made using a SONY camera model DCR-DVD202E PAL. No additional microphones other than the one in the camera was attached to the

infants' clothes or installed in the room. This had the advantage of obtaining more naturalistic data, because infants could move freely around the house and they did not play with a strange object attached to their clothes. However, it also had the disadvantage that the data did not have a perfect acoustic quality. In order to palliate this effect, the author of the recordings tried to be as close as possible to the infants without interfering with their activities.

2.2.3. Data coding

The present study analyzes infants' gesture and speech productions at ages 11, 13, 15, 17 and, 19 months of age. Infants were recorded weekly at 11 months of age and biweekly from 13 to 19 months of age. A total of 39 sessions thus yielded a total of approximately 24 h of video stream. This age range was intended to include the infants' babbling period because it is at this stage that infants produce their first communicative gestures, mostly in the form of pointing and reaching gestures. The age span included in our study therefore constitutes an earlier span than that analyzed in Butcher and Goldin-Meadow (2000), who started analyzing infants as soon as they were at the single-word period and finished their analysis when infants produced two-word combinations.[4]

[4] Due to individual differences, the infants' ages in Butcher and Goldin-Meadow (2000) varied significantly: one child was analyzed from 12 to 25 months of age, one from 13 to 19 months of age, two infants from 15 to 21 months of age, one from 15 to 25 months of age, and another one from 21 to 27 months of age.

Following Boysson-Bardies and Vihman (1991), the onset of word production was established as the first session in which the infant used four or more words (the 4-word point), whereas the first session in which approximately 25 or more words were used spontaneously was identified as the starting point for the single-word period. All four infants were at the babbling stage at 11 and 13 months of age because they were still not producing 4 words during these recording sessions, and all four infants were already at the single-word period at 17 months of age, because at that point they produced 25 or more words during a recording session. Individual differences were found at 15 months of age: at this age three infants produced around 4 words during one recording session, and the other infant produced around 20 words per session. Fifteen months of age was defined as the onset of word production given that all the infants were producing 4 words or more at this point. Table 1 summarizes the number of recorded sessions that were included in the study, classified by the infants' age, duration of the sessions, and number of words produced during them. As this table shows, we analyzed 39 recording sessions of about 30 min each, which means a total of 24 h of video recordings.

All communicative acts (visual and/or vocal) produced by the infants were identified and located in the recordings by the first author using the ELAN annotation tool (Lausberg & Sloetjes, 2009). An act was considered to be communicative if (a) the coder perceived or judged the infant's act as based on awareness and deliberate execution (Feldman & Reznick, 1996), if (b) infants produced it in a joint attention frame (either the infant directed the

gaze to the caregiver before or after the gesture, or the caregiver was attending to what the infant was doing), or if (c) the parental reactions before or after the acts suggested so. The adults' perception of the infants' acts as being intentional has been widely used in previous studies as a measure for investigating the infants' development of language and cognitive capacities (Butcher & Goldin-Meadow, 2000; Feldman & Reznick, 1996; Papaeliou & Trevarthen, 2006; Rochat, 2007). Following Iverson and Goldin-Meadow (2005), Ozçalişkan and Goldin-Meadow (2005) and So, Demir, and Goldin-Meadow (2010), we excluded from the database all hand movements that involved direct manipulation of an object or were part of a ritualized game. The approximately 24 h of recordings were thus segmented into 4,507 communicative acts, and then further classified as being 'speech-only' (*N* = 3,110), 'gesture-only' (*N* = 668), or a 'gesture-speech combination' (*N* = 729).

Participant	*Age*[5]	*Duration*	*Number of words per session*
An	0;11.03	0:33:00	0
	0;11.08	0:36:34	1
	0;11.15	0:36:35	0
	1;1.10	0:37:21	3
	1;1.24	0:41:48	2
	1;3.07	0:29:10	15
	1;3.28	0:34:49	21
	1;5.07	0:25:29	23
	1;5.28	0:33:54	26
	1;7.05	0:35:42	50
	1;7.16	0:34:21	56

[5] Years; months.days

Bi	0;11.12	0:36:20	0
	0;11.18	0:34:21	0
	0;11.25	0:26:09	0
	1;1.07	0:34:59	1
	1;1.20	0:34:05	0
	1;3.15	0:35:57	6
	1;3.29	0:35:31	9
	1;5.03	0:37:55	17
	1;5.17	0:37:58	22
	1;7.26	0:37:12	34
Ma	0;11.05	0:34:43	0
	0;11.12	0:39:44	1
	0;11.19	0:35:20	0
	0;11.25	0:33:23	1
	1;1.14	0:31:17	2
	1;1.27	0:33:36	4
	1;3.08	0:35:48	5
	1;3.22	0:32:56	7
	1;5.23	0:34:51	26
	1;7.05	0:36:29	31
On	0;11.14	0:26:25	1
	0;11.23	0:37:12	1
	1;1.06	0:37:50	2
	1;1.28	0:36:15	4
	1;3.08	0:23:28	6
	1;3.21	0:36:43	9
	1;5.15	0:37:09	22
	1;7.14	1:10:54	27
TOTAL	39 sessions	23:16:00	

Table 1. Recorded sessions included in the study, classified by infants' age, duration of the session and number of words produced per session.

To test the reliability of locating communicative acts and deciding whether they were speech-only, gesture-only, or a gesture-speech combination, two inter-transcriber reliability tests were conducted

with a subset of 10% of the data (450 cases) by two independent coders. We made sure that all infants and ages were uniformly represented. The overall agreement for the location of communicative acts was 83% and the free marginal kappa statistic obtained was 0.67, indicating that there was substantial agreement between coders regarding the identification and location of communicative acts. The overall agreement for the classification of communicative acts into one of the three categories (namely, speech-only, gesture-only, or gesture-speech combination) was 87% and the kappa statistic was of 0.81, indicating that there was almost perfect agreement between coders.

a) Speech coding

All infants' communicative acts containing speech were further annotated. First, they were annotated as containing a vocalization, if the speech sound conveyed communicative purpose but did not resemble any Catalan word, or a word, if the speech sound was clearly a Catalan word or was used consistently. This coding was used to assess the infants' lexical development. Second, all speech involving simultaneous acts was annotated to determine (a) the limits of the vocalization or word, i.e., its starting and end points (second tier in Figure 2), and (b) the limits of prosodic prominence, i.e., starting and end points of the accented syllable (first tier in Figure 2). If the accented syllable of the vocalization or word was not clearly identified, it was coded as an extra category called fuzzy

accented syllable and excluded from the statistical analyses (N = 65). Figure 2 shows an example of the acoustic labeling in Praat (Boersma & Weenink, 2012) that was later imported into ELAN and Figure 3 summarizes the speech coding conducted.

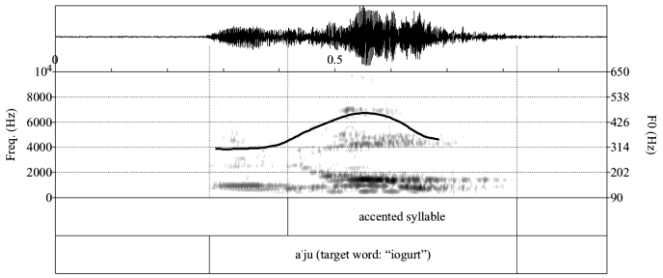

Figure 2. Example of acoustic labeling in Praat of the word [a'ju] (target word *iogurt* – 'yoghourt') produced by An at 17 months of age.

The annotation of prosodic prominence was conducted perceptually and at word-stress level. Catalan is a stress-accent language in which lexically stressed syllables generally serve as the main landing site for phrasal pitch accents (Prieto et al., 2013). Word stress always hits one of the last three syllables of the morphological word. Prieto (2006) analyzed a corpus of adults addressing infants and found that 35% of the words were monosyllables, 49% were disyllables and 13% were trisyllables. The remaining 3% of the data corresponded to longer words. Among the disyllabic forms, 63% were trochees and 37% iambs. Finally, among the trisyllabic forms, 72% were amphibracs. Importantly, no analysis of acoustic correlates of prominence such as duration or F0 tonal alignment was performed in our study because these correlates are still not stable at the ages in which

infants were analyzed (Astruc, Payne, Post, Vanrell, & Prieto, 2013; Bonsdroff & Engstrand, 2005; Engstrand & Bonsdroff, 2004; Frota & Vigário, 2008; Payne et al., 2011).

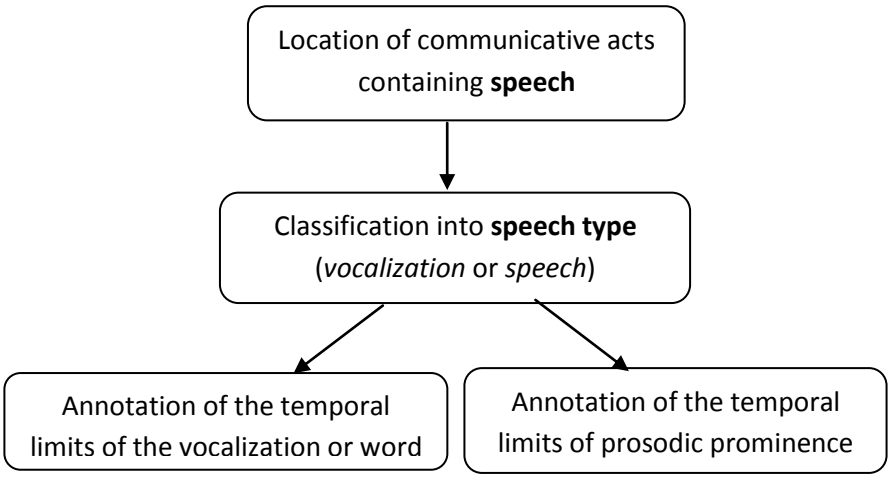

Figure 3. Summary of the steps followed during the speech coding.

b) Gesture coding

All communicative acts containing a gesture were coded using ELAN to determine their gesture type (tier 2 in Figure 4). The following categories were taken into account (following Blake, O'Rourke, & Borzellino, 1994; Capone & McGregor, 2004; Ekman & Friesen, 1969; Iverson, Tencer, Lany, & Goldin-Meadow, 2000): pointing gesture, a deictic gesture infants produce when extending the arm and the index finger towards an entity in order to direct the caregiver's attention to it; reaching gesture, a deictic gesture produced when the infant extends the arm and opens the hand

towards an entity in order to direct the caregiver's attention to it; conventional gesture, ritual actions such as head nodding to mean 'yes', head shaking to mean 'no', bye-bye gesture, clapping hands, kissing gesture, 'sh' gesture, and negating with the index-finger extended; emotive gesture, the infant's expression of an emotional state, such as shaking arms when being angry, or shaking legs to protest, as opposed to the transmission of information; and finally other gestures, when the infant produced a proto-beat gesture, or an object-related action resembling an iconic gesture.

Next, all gestures classified as either pointing or reaching (i.e., which shared the feature of being deictic) were annotated regarding their motivation or intentionality. Gesture motivation was annotated in order to investigate potential influences of this factor on the temporal alignment of gesture and speech. Two categories were taken into account in this annotation, imperative or declarative (tier 3 in Figure 4). A deictic gesture had an imperative motive if infants used it to ask the adult to retrieve an object for them, and a declarative motive if infants used it to share attention or inform the adult about something. Most of the studies on pointing development support the dichotomy between imperative and declarative pointing gestures that was first proposed by Bates et al. (1975) and later corroborated by further research (Camaioni et al., 2004; Cochet & Vauclair, 2010; Liszkowski, 2007; Tomasello et al., 2007)[6].

[6] The distinction between imperative and declarative has alternatives in the literature: Begus and Southgate, (2012) and Southgate, van Maanen, and Csibra (2007) propose that all infant pointing gestures have an interrogative function,

To test the reliability of the gesture coding, the two independent coders that also participated in the previous reliability tests conducted two inter-transcriber reliability tests with a random subset of 20% of the data (145 cases), one in terms of gesture type and another in terms of gesture motive, in which all infants and all ages were uniformly represented. In terms of gesture type (pointing, reaching, emotive, conventional, or others), overall agreement between coders was 95% and the kappa statistic value was 0.94, suggesting almost perfect agreement. For gesture motive (imperative or declarative), overall agreement was 86% and the kappa statistic value was 0.73, suggesting again a high degree of agreement between coders.

Finally, all gesture-speech combinations containing a pointing or reaching gesture were annotated in terms of their gesture phases (tier 4 in Figure 4), following McNeill's (1992) and Kendon's (2004) observational measures: (a) the preparation phase, in which the arm moves from rest position until the stroke of the gesture; (b) the stroke phase, the interval of peak of effort in the gesture that expresses the meaning of the gesture (McNeill, 1992:83) or, in other words, the phase when the 'expression' of the gesture, whatever it may be, is accomplished and in which the movement dynamics of 'effort' and 'shape' are expressed with greater clarity (Kendon, 2004:112) (c) the retraction phase, in which the arm moves from the stroke position to rest position again. Additionally, another measure

---

Leavens (2009) state that they all have an instrumental function, and Moore and D'Entremont (2001) argue that all pointing gestures are motivated egocentrically.

was annotated within the stroke of the gesture, namely the gesture apex (tier 5 in Figure 4). Whereas the stroke of the gesture is an interval of time in the case of a deictic gesture during which the arm is maximally extended, the gesture apex is the specific point within the stroke interval at which the finger is maximally extended.

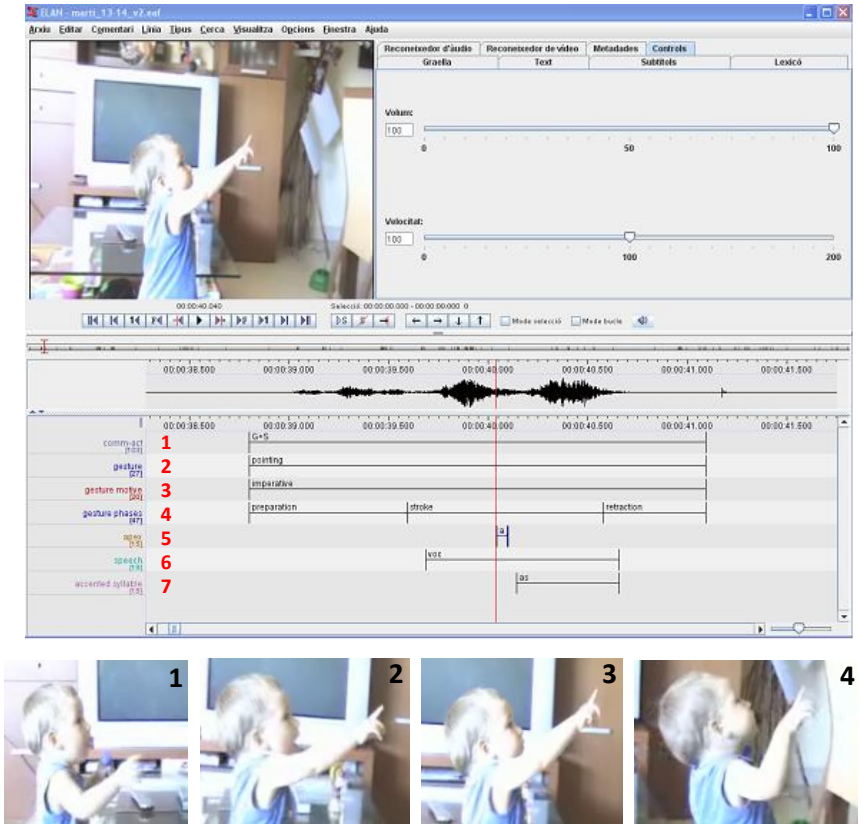

Figure 4. ELAN still image with all the annotated tiers (from 1 to 7). Lower frames, four specific enlarged images taken in the course of a pointing gesture which illustrate the four phases of a pointing gesture: (1) the preparation phase, (2) the stroke phase and before the apex is reached, (3) the apex, and (4) the retraction phase.

In order to locate the stroke and apex of the pointing gesture, we examined the video file (following Esteve-Gibert & Prieto, 2013; Levelt et al., 1985; Rusiewicz, 2010). ELAN allows precise navigation through the video recording, i.e., frame by frame. Though the software program can in principle permit an even more precise annotation (2 ms by 2 ms), this option could not be applied because the video was recorded at a frame rate of 25 frames per second. First, the stroke of the gesture was annotated in those video frames in which the arm was well extended with no blurring of the image, the fingertip being fully extended or not. Despite the absence of image blurring, the arm was not totally static during the interval of the gesture stroke, with the fingertip moving a few pixels back and forth. Next, the gesture apex was annotated in the specific video frame in which we located the furthest spatial excursion of the fingertip during the interval of time in which the arm was maximally extended (see still images at the bottom of Figure 4). When infants performed the pointing gesture more slowly, this gesture peak could last more than one frame (normally two frames). In such cases, the gesture peak was considered to be the last of these video frames. Figure 5 summarizes the procedure followed for gesture coding and shows an example of each type of gesture.

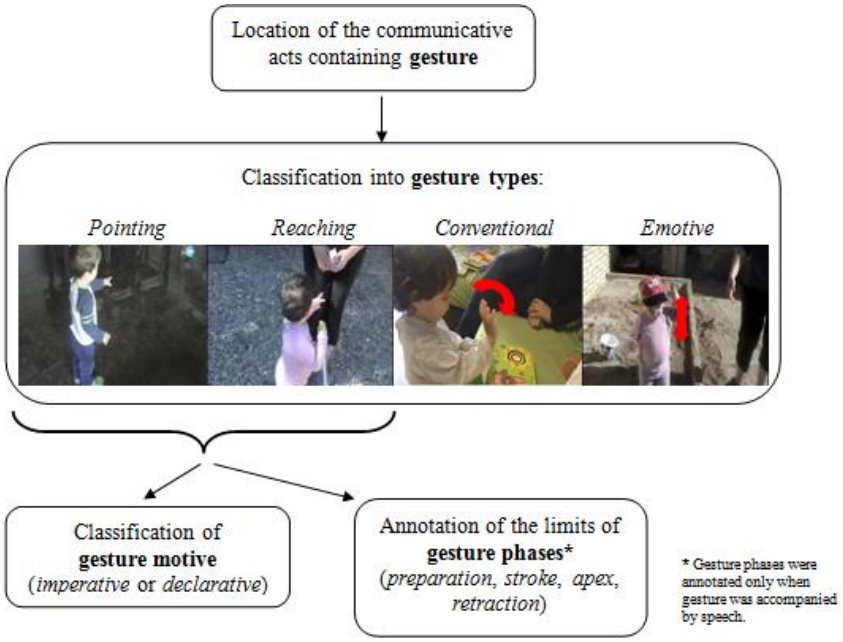

Figure 5. Summary of the gesture coding with a still image of every gesture type taken into account.

2.3. Results

The main goal of this study was to analyze the early development of gesture and speech patterns, and it can be divided in two specific goals: (1) how do infants combine gesture with speech across ages, and (2) how do they temporally align the two modalities across age groups. Results are presented in two main sections, one for each aim.

2.3.1. How do infants combine gesture with speech across ages?

In this section we explore how infants combine gesture and speech across ages. Three main questions will be addressed: (1) When do infants start producing most of their gestures in combination with speech? (2) In the first gesture-speech combinations, which gesture types do infants produce? And (3) which intentions do infants convey in their first pointing gesture-speech combinations?

We first examined when infants start combining gestures with speech. Figure 6 shows the distribution of 'gesture-only' and 'gesture-speech combination' acts produced by infants across ages (Table 2 shows the raw numbers, including also the speech-only group). Of all the communicative acts containing gestures, at 11 months of age most do not yet involve speech. Thirteen-month-old infants produce roughly the same number of gestures accompanied by speech and gestures without speech, and from 15 months onwards the proportion of gesture-speech combinations is higher than the proportion of gesture-only acts. Chi-squared tests of independence tested the ratio of 'gesture-only' to 'gesture-speech combination' acts across age groups. Results showed that the ratio of 'gesture-only' to 'gesture-speech combination' was statistically different at 11 months (χ^2 (1, N = 455) = 24.231, p < .001), at 15 months (χ^2 (1, N = 268) = 20.433, p < .001) and at 19 months (χ^2 (1, N = 166) = 38.554, p < .001), but not at 13 (χ^2 (1, N = 259) = .004, p = .950) nor at 17 months (χ^2 (1, N = 249) = .486, p = .486). These results indicate that 11-month-old infants produced most of their

gestures without accompanying speech, and that it is not until infants were 15 months of age that this tendency changes to such an extent that most of their gestures were produced together with speech, as seen in adults.

Age (in months)	*Speech-only*	*Gesture-only*	*Gesture-speech comb.*	*Total*
11	710	280	175	1,165
13	576	129	130	835
15	722	97	171	990
17	556	119	130	805
19	546	43	123	712
Total	3,110	668	729	4,507

Table 2. Description of the data included in the analysis as a function of the ages and communicative act types.

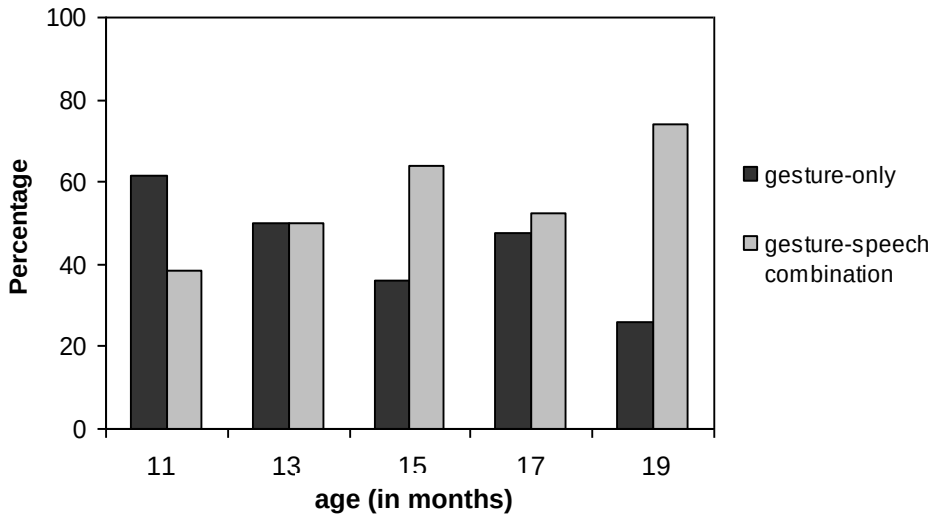

Figure 6. Ratio (expressed in percentages) of the distribution of all communicative acts containing gestures into the category of 'gesture-only' or 'gesture-speech combination'.

With regard to the gesture types that infants produced when combining gesture with speech, Table 3 shows the distribution of gesture types across age groups in the gesture-speech combinations. The results show that deictic gestures were the most frequent gestures across all age groups. At 11 months of age most of the gestures infants produced were deictic gestures (48.6%, divided into 34.9% pointing and 13.7% reaching), and emotive gestures (30.9%). At 13 months the proportion of deictic gestures increased to 56.9% (40% pointing and 16.9% reaching) and emotive gestures represented 29.3% of the gestures. Fifteen-month-old infants produced more deictic gestures than any other type of gesture (53.8% pointing and 20.5% reaching). At 17 months of age conventional gestures increased compared to the previous ages but most gestures were nonetheless still deictic (40.8% pointing and 13.8% reaching). Finally, at 19 months the proportion of pointing deictic gestures was higher than at all previous ages (65% pointing and 13.8% reaching).

Months	*Pointing*		*Reaching*		*Conventional*		*Emotive*		*Other*	
	N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
11	61	34.9	24	13.7	16	9.1	54	30.9	20	11.4
13	52	40.0	22	16.9	9	6.9	38	29.3	9	6.9
15	92	53.8	35	20.5	20	11.7	16	9.4	8	4.6
17	53	40.8	18	13.8	37	28.6	14	10.8	8	6.1
19	80	65.0	17	13.8	17	13.8	8	6.5	1	0.8

Table 3. Total numbers and percentages of gesture types across ages in gesture-speech combinations.

Regarding the gesture motives behind the deictic gestures, Figure 7 (top panel) shows that at all infants produced a higher proportion of deictic gestures with a declarative than imperative intention. A look at the gesture types that infants used to convey these distinct gesture motives shows that the imperative motive was conveyed mostly by reaching gestures and the declarative intention was conveyed by pointing gestures (Figure 7, bottom panels).

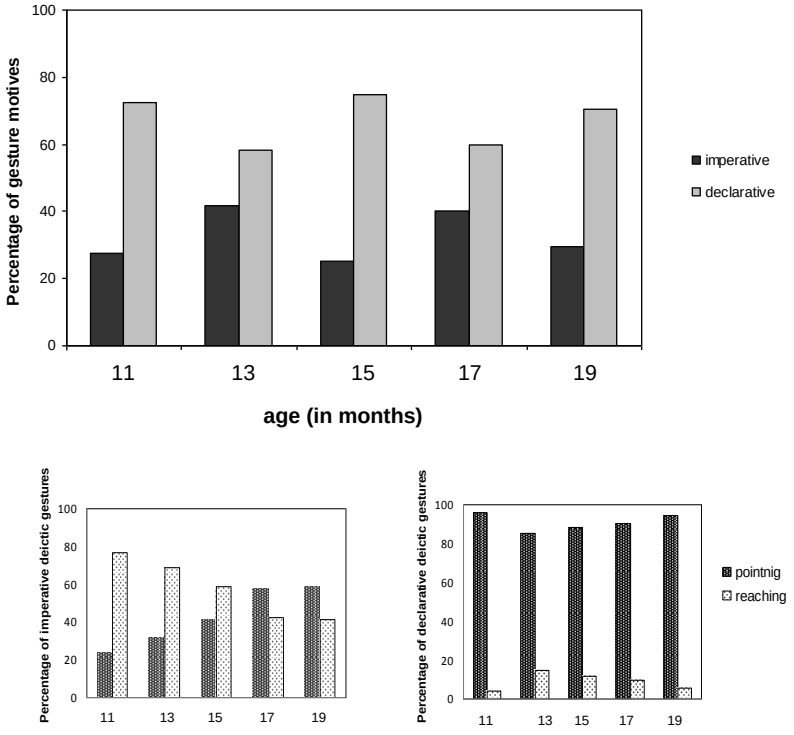

Figure 7. Relative proportions of imperative and declarative gestures across ages (top panel). Bottom panels, ratios of gesture types within imperative deictic gestures (left panel) and declarative deictic gestures (right panel).

The intention of deictic gestures was analyzed to investigate potential influences of this factor on the duration of the gesture and consequently on the temporal alignment of early gesture-speech combinations. Our hypothesis was that imperative pointing and reaching gestures would be longer, and thus would overlap more and be more easily coordinated with speech. However, the statistical analyses revealed that no such effect occurred. Specifically, LMM analyses were carried out with stroke duration as the dependent variable, gesture motive (2 levels: imperative, declarative) as the fixed factor, and subject as a random factor. Results revealed no main effect of gesture motive on the duration of the gesture stroke ($F(1,4.47) = 1.453$, $p = .229$). For this reason gesture motive was not included as a fixed factor in any of the subsequent analyses of the temporal alignment of different gesture and speech landmarks.

2.3.2. How do infants temporally align gesture and speech across ages?

The main goal of this section is to assess how infants temporally align deictic gesture and speech combinations across the ages analyzed. Following the adult studies on gesture-speech temporal alignment, the alignment between deictic gesture-speech combinations was analyzed at four different levels: (1) the temporal relationship between the gesture onset and the onset of the vocalization, to compare our results in infants with studies for

adults suggesting that the onset of the gesture always precedes the onset of speech (Butterworth & Beattie, 1978; Ferré, 2010; Levelt et al., 1985); (2) the temporal alignment between the stroke onset and the onset of speech, to compare our results in infants with results for adults suggesting that the stroke of the gesture aligns with the onset of speech (Bergmann et al., 2011; Ferré, 2010; Roustan & Dohen, 2010); (3) the alignment between the stroke onset and the onset of the accented syllable in speech, to compare our results with results for adults suggesting that gesture strokes are aligned with prosodically prominent syllables (Krahmer & Swerts, 2007; Leonard & Cummins, 2010; Loehr, 2012); and (4) the alignment between the gesture apex and the end of the accented syllable in speech, to compare our results in infants with studies on adult alignment suggesting that the gesture apex occurs within the limits of the prosodically prominent syllables (De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Rochet-Capellan et al., 2008).

All statistical analyses in this section were performed by applying a linear mixed model (LMM; West, Welch, & Galecki, 2007) using SPSS Statistics 16.0 (SPSS Inc., Chicago IL). West et al., 2007), and Baayen, Davidson, and Bates (2008) state that LMMs are the appropriate model for analyzing unbalanced longitudinal data, since they allow for subjects with missing time points (i.e., unequal measurements over time for individuals), have the capacity to include all observations available or all individuals in the analysis and cope with missing data at random. As the authors point out, linear mixed models can accommodate all of the data that are

available for a given subject, without dropping any of the data collected from that subject.

First, the alignment between the gesture onset and the onset of the associated speech was analyzed. The temporal distance between the onset of the vocalization and the gesture onset was the dependent variable, age was the fixed factor (5 levels: 11, 13, 15, 17 and 19 months of age), and subject was a random factor. The analysis revealed a statistically significant effect of age on the distance between gesture onset and onset of speech ($F(4,462.530) = 9.998$, $p < .001$) (see Table 4 for statistic coefficients). LSD pair-wise comparisons revealed that the mean distance between gesture onset and onset of speech varied significantly between 11 and 13 months of age ($p < .05$), between 13 and 15 months of age ($p < .001$), between 11 and 17 months of age ($p < .001$), between 11 and 19 months of age ($p < .001$), between 13 and 19 months of age ($p < .01$), and between 15 and 19 months of age ($p < .05$). Figure 8 shows that the gesture onset preceded the onset of speech at all ages, but that the tendency is for this distance to decrease as the infant grows up and for the variance in this distance measure to decrease as well. In adult studies, it has been found that gesture onset precedes speech onset (Butterworth & Beattie, 1978; Ferré, 2010). None of these studies, however, detail the time lag they found between the onset of gesture and the onset of speech. Our results thus show that infants align these two landmarks in an adult-like way in the sense that gesture onset occurs before speech onset.

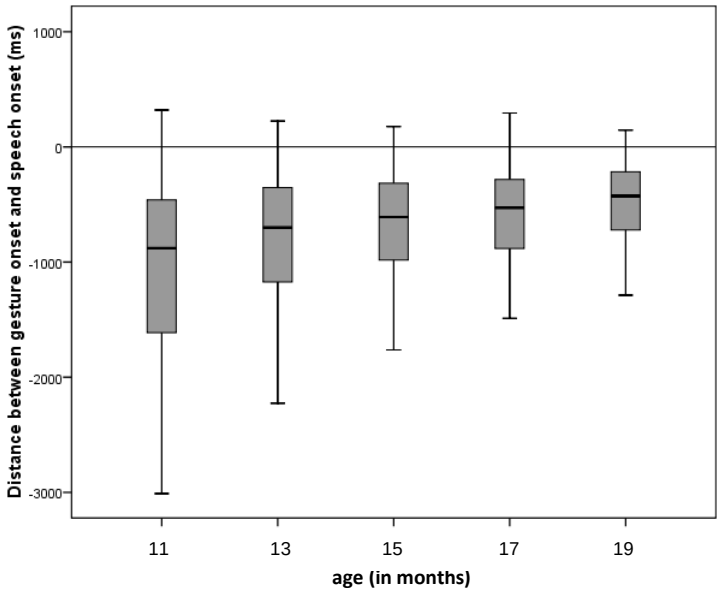

Figure 8. Distance between gesture onset and the onset of speech across ages (in milliseconds). Positive values (> 0) represent those cases in which the gesture onset occurs after the onset of speech, while negative values (< 0) represent cases in which the gesture onset occurs before the onset of speech.

Dependent variable: Distance between gesture onset and onset of speech (in ms)					
	Estimates	*Residual variance*	*Standard deviation*	*F value*	*P value*
Fixed factor (age)	-505.53	-	756.88	9.998	.000***
Random factor (subject)	-	523063.03	723.23	-	-

Table 4. Coefficients of the distance between gesture onset and onset of speech (in ms), with estimates, standard deviation, *F* value and *p* value for the fixed factor, and residual variance and standard deviation for the random factor.

Second, we introduced a more fine-grained coordination measure which takes into account the temporal position of the most prominent period in the deictic gesture, namely the stroke (following Bergmann et al., 2011; Ferré, 2010; Roustan & Dohen, 2010). The temporal distance between the onset of the stroke and the onset of speech was the dependent variable, age was the fixed factor (5 levels: 11, 13, 15, 17 and 19 months), and subject was included as a random factor. The statistical analysis indicated that age did not have an effect on this distance ($F(4,434.639) = 2.066$, $p = .084$) (see Table 5 for statistic coefficients). Figure 9 displays the distance between stroke onset and onset of speech across ages. The figure shows a close temporal alignment between the two modalities across ages, as found in adult studies. Though the age factor was not significant, a slight tendency is observed in Figure 9 during the babbling stage (11 and 13 months old), the stroke onset was aligned with the onset of speech, and as the infants' linguistic abilities developed (15 month onwards), their speech tended to start slightly before the stroke. Studies in adult gesture-speech alignment which took into account these two measures found that 72% of the gesture strokes start before the onset of speech (Ferré, 2010), that stroke onset precedes speech onset on average by 129.89 ms (Bergmann et al., 2011) and that the maximum extension of the finger occurs within or close to the focus constituent (Roustan & Dohen, 2010).

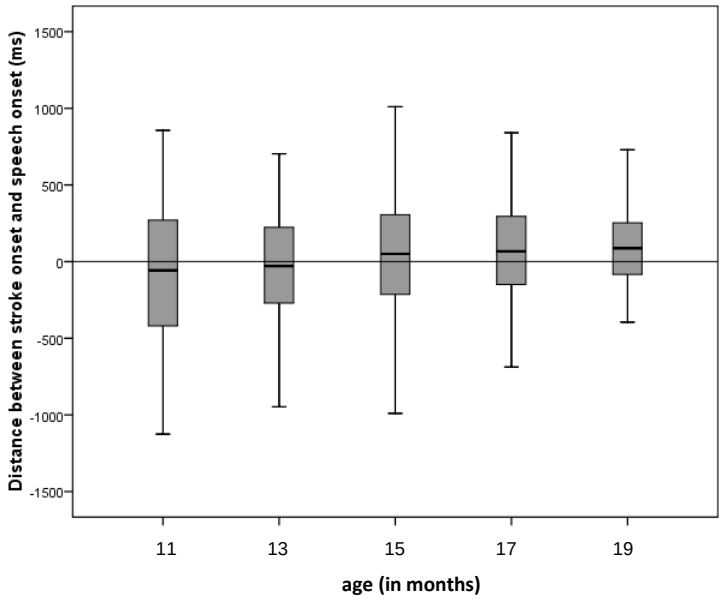

Figure 9. Distance between stroke onset and onset of speech across ages (in milliseconds). Positive values (> 0) represent those cases in which stroke onset occurs after speech onset, while negative values (< 0) represent cases in which stroke onset occurs before speech onset.

Dependent variable: Distance between stroke onset and onset of speech (in ms)					
	Estimates	*Residual variance*	*Standard deviation*	*F value*	*P value*
Fixed factor (age)	79.31	-	3123.93	2.066	.084
Random factor (subject)	-	157430.99	396.77	-	-

Table 5. Coefficients of the distance between stroke onset and onset of speech (in ms), with estimates, standard deviation, *F* value and *p* value for the fixed factor, and residual variance and standard deviation for the random factor.

Third, we introduced the location of the prosodically prominent syllables in the coordination analysis, as adult studies have highlighted the importance of this anchoring site (Krahmer & Swerts, 2007; Leonard & Cummins, 2010; Loehr 2012). The alignment between the stroke onset and the onset of accented syllable was analyzed. We expected an even tighter alignment between the start of the gesture stroke and the start of the accented syllable as infants develop, since studies with adults have reported a close alignment between these two landmarks. Thus, the dependent variable was the distance between stroke onset and the onset of the accented syllable, the fixed factor was age (5 levels: 11, 13, 15, 17 and 19 months), and subject was introduced as a random factor. The statistical analysis indicated that age did not have an effect on the distance between the stroke onset and the onset of accented syllable ($F(4,431.213) = .595$, $p = .667$) (see Table 6 for statistic coefficients). Figure 10 displays the distance between onset of the stroke and the onset of the accented syllable across ages. Though age did not significantly affect this measure, a tendency can be observed: at the babbling stage (at 11 and 13 months of age), the stroke tended to slightly precede the onset of the accented syllable, whereas from the onset of word production onwards (from 15 to 19 months of age), the stroke onset and the onset of the accented syllable were very closely aligned, and less data variance was observed. These results are in accordance with adult studies showing a co-occurrence between those landmarks (Krahmer & Swerts, 2007; Leonard & Cummins, 2010; Loehr 2007).

Dependent variable: Distance between stroke onset and accented syllable onset (in ms)					
	Estimates	*Residual variance*	*Standard deviation*	*F value*	*P value*
Fixed factor (age)	-1.004	-	412.383	.595	.667
Random factor (subject)	-	168135.28	410.043	-	-

Table 6. Coefficients of the distance between stroke onset and accented syllable onset (in ms), with estimates, standard deviation, *F* value and *p* value for the fixed factor, and residual variance and standard deviation for the random factor.

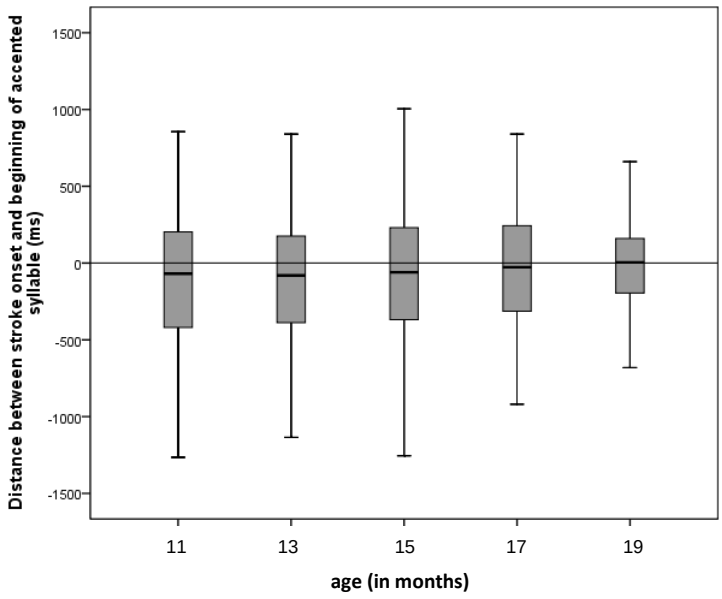

Figure 10. Distance between stroke onset and onset of the accented syllable across ages (in milliseconds). Positive values (> 0) represent those cases in which stroke onset occurs after onset of accented syllable, while negative values (< 0) represent cases in which stroke onset occurs before onset of accented syllable.

Fourth, the temporal distance between the gesture apex and the end of the accented syllable was also analyzed. The dependent variable was the distance between the gesture apex and the end of the accented syllable, the fixed factor was age (5 levels: 11, 13, 15, 17 and 19 months), and subject was the random factor. The statistical analysis showed that age did not have a main effect on the position of the gesture apex within the accented syllable ($F(4,439.933) = 1.127$, $p = .343$) (see Table 7 for statistic coefficients). Thus, at all ages the gesture apex preceded the end of the accented syllable (see Figure 11), but different tendencies can be located further from the end of the accented syllable, and the box plots show higher variation; as the infants' linguistic abilities developed, however, the gesture apex tended to occur closer to the end of the accented syllable and the variation diminished. Adult studies on the alignment between gesture apex and the accented syllable showed that the gesture apex occurs between 350 and 0 ms prior to the end of the accented syllable (De Ruiter, 1998; Esteve-Gibert & Prieto, 2013), so our results show that infants align these two landmarks in an adult-like way at the babbling stage in the sense that they produce the gesture apex before the accented syllable is finished.

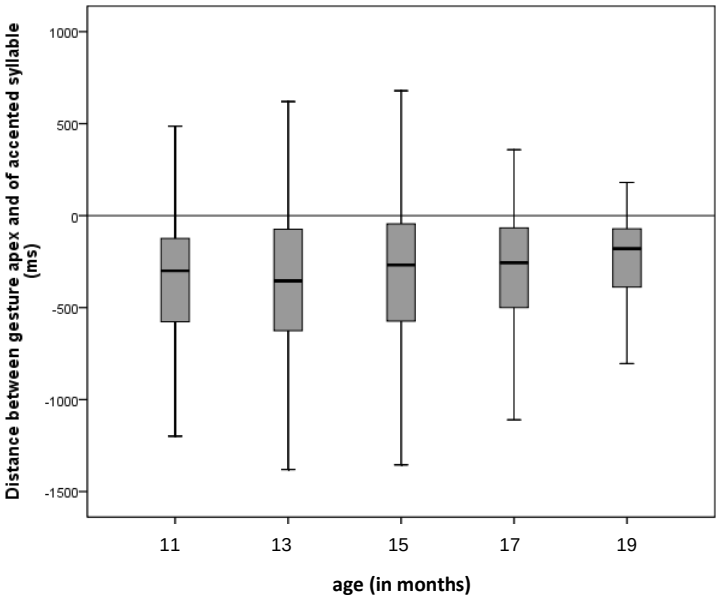

Figure 11. Distance between the gesture apex and the end of the accented syllable across ages and as a function of the metrical patterns (in milliseconds). Positive values (> 0) represent those cases in which the gesture apex occurs after the end of the accented syllable, and negative values (< 0) are those cases in which the gesture apex occurs before the end of the accented syllable.

Dependent variable: Distance between gesture apex and end of accented syllable (in ms)					
	Estimates	*Residual variance*	*Standard deviation*	*F value*	*P value*
Fixed factor (age)	-221.52	-	402.752	1.127	.343
Random factor (subject)	-	153665.47	392.002	-	-

Table 7. Coefficients of the distance between gesture apex and end of accented syllable (in ms), with estimates, standard deviation, *F* and *p* values for the fixed factor, and residual variance and standard deviation for the random factor.

Summarizing, infants show an adult-like pattern of coordination between the distinct gesture and prosodic landmarks analyzed across ages. First, gesture onset precedes speech onset (Figure 8), although there are significant differences across ages: at the babbling stage, the distance between the two points is higher and there is more variation in the data, whereas at the single-word period this distance is more similar to what previous studies have found in adult data, and the variation is significantly reduced. Second, gesture and speech are very tightly aligned when landmarks take into account prominence: infants align more closely the beginning of the gesture stroke with the speech onset, and also the beginning of the stroke with the beginning of the accented syllable. Importantly, age was not a significant main factor in either of the two coordination analyses, yet Figures 9 and 10 show that the absolute distance in time and variance across age groups is smaller in older infants. Third, the temporal distance between gesture apex and the end of the accented syllable shows adult-like patterns in the sense that this gesture apex precedes the end of the accented syllable (Figure 11). Although this alignment does not vary significantly across ages, a finer alignment is again observed at the single-word period. Altogether, these findings suggest clear adult-like patterns of gesture-speech integration in the very first multimodal utterances produced by the infants already at the babbling stage.

2.4. Discussion

This study explored the patterns of gesture and speech combinations in infants from the babbling to the single-word period, as well as the temporal alignment between the two modalities when they are combined. The analysis of approximately 24 h of naturalistic recordings of four Catalan infants in five consecutive developmental stages, namely at 11, 13, 15, 17 and 19 months of age, provided us a total of 4,507 communicative acts. An infant's act was considered to be intentional if (a) the coder perceived or judged the infant's act as based on awareness and deliberate execution, if (b) infants produced it in a joint attention frame, or if (c) the parental reactions before or after the acts suggested so. While these measures are not totally objective, they have been proven to be a reliable measure when correlating the adults' inclination to interpret infants' acts as intentional and the infants' later development of cognitive capacities (Olson, Bates, & Bayles, 1982; Sperry & Symons, 2003). In the present study, two independent coders performed an inter-transcriber reliability test by identifying the communicative acts from 10% of the observational data. Results of this analysis reflect that although this method resulted into some coding errors, there was still substantial agreement between coders when locating communicative acts (the overall agreement was 83% and the free marginal kappa statistic was 0.67).

Summarizing, three main results can be highlighted from the data: first, it is from the onset of word production that gesture starts to be

produced mainly in combination with speech with an intentional purpose; second, in these early gesture-speech combinations most of the gestures are deictic gestures (pointing and reaching) with a declarative communicative purpose; and third, there is clear evidence of temporal alignment between deictic gesture and speech already at the babbling stage in the sense that (a) gesture onset starts before speech onset, (b) the stroke onset is temporally aligned with the onset of speech and with the onset of the prominent syllable, and (c) the gesture apex occurs before the end of the accented syllable. In the following paragraphs we discuss one by one these results in more detail, reporting the main findings suggested by our statistical analyses. Although only four infants were analyzed in this study, we believe that the large amount of data obtained can compensate for the small number of subjects.

The results of the longitudinal analysis show that it is not until the onset of word production that infants combine communicative gesture and speech significantly more often than producing gesture-only acts. Our results show that 11-month-old infants produce more gesture-only acts than gesture-speech combinations. At 13 months of age the proportion of gesture- speech combinations is still not higher than the proportion of gesture-only acts, though they have increased with respect to the previous ages analyzed. However, from 15 months onwards infants start producing their first words and also start combining gesture with speech significantly more often than producing gesture-only acts. These results confirm those by Butcher and Goldin-Meadow (2000). The authors started their analysis when infants began producing their first words and ended it

when they produced two-word combinations and found that infants started combining gesture with speech at the transition between the one-word and two-word periods. Our study has enlarged the window analyzed by Butcher and Goldin-Meadow (2000), by focusing on the development of multimodality at the transition from the babbling period to the single-word period. All in all, our findings support those of Butcher and Goldin-Meadow (2000) with respect to an early infants' ability to combine gesture and speech.

A second finding of the study is that deictic gestures (pointing and reaching) are the most frequent gesture-speech combinations in the age range analyzed, and that at all ages infants produced more deictic gestures with a declarative purpose than with an imperative purpose. Interestingly, we observe that imperative deictic gestures mostly take the form of reaching gestures, whereas declarative deictic gestures almost always take the form of pointing gestures, corroborating previous studies in the field (Cochet & Vauclair, 2010; Colonnesi et al., 2010; Leung & Rheingold, 1981). However, our results could be biased by the specific contexts in which these infants were recorded, as they were recorded during free-play sessions at their homes and while interacting with their parents, and these play situations might not reflect the total output of gestures infants produce. For instance, it might well be that in an eating situation infants produce a higher proportion of imperative pointing gestures compared to declarative ones. Importantly, it should be borne in mind that our analysis of the development of the gesture motives and gesture types constitutes a description of the corpus under analysis and no strong conclusions can be drawn from it.

An important focus of our investigation was to describe developmental patterns related to how early gesture and speech combinations are temporally aligned. In order to investigate this, we analyzed in detail the temporal alignment between gesture and speech by taking into account specific measurements that previous studies with adults had proposed to be crucial in characterizing gesture- speech alignment. Studies in adult multimodal communication have shown that gesture onset precedes speech onset (Butterworth & Beattie, 1978; Ferré, 2010; Levelt et al., 1985), that the gesture stroke slightly precedes the onset of speech (Bergmann et al., 2011; Ferré, 2010; Roustan & Dohen, 2010), that the most prominent part of the gesture coincides with the most prominent part of speech (Krahmer & Swerts, 2007; Leonard & Cummins, 2010; Loehr, 2012), and that the gesture apex is produced within the limits of the prominent syllable in speech (De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Rochet-Capellan et al., 2008). Our analysis reveals that infants' behavior shows all these alignments even before they can produce their first words.

First, our results on the distance between gesture onset and onset of speech reveal that 11-month-old infants show the adult-like pattern in that gesture starts before speech. However, our statistical analysis also revealed that age significantly affects this measure because in the babbling period the distance between the two measures is significantly higher than in the single-word period.

Second, our results show that already at 11 months of age and across all the stages analyzed infants align the gesture stroke with

the onset of speech because they produce these two landmarks simultaneously. Some studies with adults have shown that the stroke onset precedes the speech onset (Bergmann et al., 2011; Ferré, 2010; Roustan & Dohen, 2010), while some others found that the gesture stroke starts when the accented syllable has already been initiated in monosyllables and trochees, and that both landmarks occur almost simultaneously in the iambic condition (Esteve-Gibert & Prieto, 2013). These different findings with adults could be due to the fact that they analyzed different types of gesture: those who found that the gesture stroke precedes the speech onset had analyzed a mixture of iconic, deictic, and discourse gestures, and those who found that the gesture stroke follows the speech onset had analyzed only deictic gestures. In the present study, infants produce much simpler speech structures at the ages analyzed, mostly monosyllables and a few disyllables, and this might influence the alignment between stroke onset and onset of speech. Importantly, our statistical analysis revealed no effect of age on the distance between stroke onset and onset of speech, contrary to what Butcher and Goldin-Meadow (2000) found in their study. These authors found that the alignment between these two measures was significantly affected by age, adult-like alignment being present at the two-word stage but not at the single-word period. The difference in results might be due to the specific anchoring points taken into account: Butcher and Goldin-Meadow (2000) considered gesture-speech combinations to be synchronous if the vocalization occurred on the stroke of the gesture or at the peak of the gesture, whereas in

the present study we took into account the onset of both the gesture stroke and speech.

Third, the tightest adult-like alignment between gesture and speech is observed when prominence in gesture and prominence in speech are taken into account: infants produce the stroke onset coinciding with the onset of the accented syllable, just as adults do. Crucially, our statistical analysis revealed that this finding is not significantly affected by age. It is interesting to note that infants temporally align the gesture stroke as closely with the onset of the accented syllable as they do with the speech onset. This fact might be because these two measures in speech are very close to each other given that most of the infants' first vocalizations and words are monosyllables or have word-initial stress.

Fourth, our analysis of the alignment between gesture apex and the end of the accented syllable revealed adult-like patterns already at the babbling stage to the extent that infants produce the gesture apex before the end of the accented syllable. Studies on adult alignment of gesture and speech have shown that gesture apex occurs between 350 ms and 0 ms prior to the end of the accented syllable (De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Rochet-Capellan et al., 2008), and our results demonstrate that infants already show signs of the same behavior before they produce their first words.

All in all, our results show adult-like patterns when infants combine and align gesture and speech at an early age in language development and before they are able to produce two-word

combinations. Thus, already at the babbling stage infants produce the gesture onset before the onset of the corresponding speech (and this alignment is finely tuned at the single-word stage), they temporally align the gesture stroke, i.e., the most prominent part of the gesture, with the speech onset and, crucially, with the onset of the accented syllable, and they produce the gesture apex before the end of the accented syllable.

These results expand on those reported in Butcher and Goldin-Meadow (2000) in two ways: first, our study focuses on infants at the transition from the babbling to the single-word period and not at the transition between one- and two- word combinations, because it is at this earlier age that infants start producing gestures like pointing or reaching with a communicative purpose; and second, our study makes a detailed analysis of the temporal alignment between the two modalities based on the latest results in the field. Specifically, three main acquisition results can be highlighted in three points: (1) infants start producing most of their gestures in combination with speech in an adult-like fashion at the early single-word period; (2) in these early combinations gesture onset always precedes speech onset, and this alignment is finely tuned at the single-word period compared to the babbling stage; and (3) both modalities are temporally aligned in an adult-like way when gesture and acoustic prominences are taken into account: the stroke onset co-occurs with speech onset at all ages, the stroke onset precedes the beginning of the accented syllable at all ages, and the gesture apex is located before the end of the accented syllable at all ages.

These results suggest that infants align gesture and prosodic structures already before they produce their first words.

2.5. Conclusion

This study was intended to contribute to the body of research on the development of gesture and speech combinations in infants. Previous studies based on the dynamic systems theory (Iverson & Thelen, 1999) suggest that the early coordination between rhythmic motor and vocal movements is a precursor to the adult system in which gesture and speech are combined (Ejiri & Masataka, 2001; Iverson & Fagan, 2004). However, few studies have investigated the specific patterns of the early coordination of communicative gestures with speech (nor are there many on early rhythmic movements). To our knowledge, only Butcher and Goldin-Meadow (2000) have previously explored the question of when infants learn to combine the two modalities and the way they align them. They analyzed infants at the transition between the one and two-word stages and found that infants started combining the two modalities at the single-word period and started aligning them in an adult way at the two-word stage. Our results on the transition between the babbling stage and the single-word period confirm those by Butcher and Goldin-Meadow (2000) in the sense that it is at the single-word period that both modalities start being combined. Also, our study extends the work by Butcher and Goldin-Meadow (2000) because we analyze infants already from the babbling stage and because we

examine in more detail the temporal alignment of deictic gesture-speech combinations.

In this respect, and based on recent findings in the literature on adult temporal alignment of gesture and speech, our analyses include coordination measurements related to prosodic and gestural prominence that have been found to play a crucial role in the temporal alignment between gesture and speech. And our results show that there is evidence of temporal alignment between communicative gesture and speech already at the babbling stage because gestures start before their corresponding speech, because the stroke onset coincides with the onset of the prominent syllable in speech, and because the gesture apex is produced before the end of the accented syllable.

Various models of gesture and speech production have investigated the relation between gestures and speech. Theoretical models of gesture production such as the 'Growth Point Theory' by McNeill (1992), the 'Sketch Model' (De Ruiter, 2000), the 'Lexical Access Hypothesis' by Krauss, Chen and Chawla (2006), or the 'Information Packaging Hypothesis' by Kita (2000) all try to account for the strong interrelation and influence between gesture and speech that characterize human communication. These models differ significantly regarding the semantic role of gestures with respect to speech and vice versa, or the phases in which gestures are conceptualized, planned, and executed. However, they all agree on the close temporal integration of gesture and speech in production.

We believe that the present findings provide evidence for this integration, and from a developmental point of view.

Indeed, our results suggest that there is a temporal alignment of communicative gesture and speech from the very first stages of language production. Yet strong claims about these integration patterns can only be made after more data is analyzed. Our study has been limited to the analysis of 4,507 longitudinal observations of naturalistic interactions between four infants and their caregivers. We think that a larger number of subjects should be analyzed in the future and that more experimental data in a controlled setting (with higher audio quality and movement trackers) will be useful to provide more solid confirmation for our claim that gesture and speech form a temporally integrated system from the onset of language production.

CHAPTER 3: NINE-MONTH-OLD INFANTS ARE SENSITIVE TO THE TEMPORAL ALIGNMENT OF PROSODIC AND GESTURE PROMINENCES

3.1. Introduction

When humans communicate we use multimodal cues (i.e., speech and gestures) that increase the efficiency of the information we transmit. There is a broad consensus in the literature that both modalities are integrated semantically and temporally in human communication (e.g., McNeill, 1992, Kelly et al., 2010). Researchers have highlighted the tight temporal relationship between gesture and speech. Typically, the most prominent part of co-speech gestures (i.e., the interval of the gesture stroke or the specific apex within the gesture stroke) co-occurs with the prosodically prominent part of speech (i.e., the accented syllable and the pitch peak within the accented syllable) (e.g., De Ruiter, 1998; Kendon, 1980; Levelt et al., 1985; McNeill, 1992). This temporal alignment is also evidenced by the fact that gesture and prosodic timing influence each other: the perception and production of prosodic prominence are both affected by the presence of an accompanying interval of gesture prominence (Krahmer & Swerts, 2007) and the prosodic structure influences the timing of the gesture movement (Esteve-Gibert & Prieto, 2013; Loehr, 2012).

Some studies have focused on how infants learn to integrate communicative gestures with speech temporally and pragmatically. The general ability to produce communicative gestures starts with

pointing gestures during the second half of the first year of life (Bates et al., 1975). After the first year of life infants already produce pointing gestures with distinct social intentions (imperative, expressive, or informative) and they are also able to comprehend the social intentions behind them (Aureli et al., 2009; Behne et al., 2012; Liszkowski, 2005). Besides, it is not until around 15 months of age that infants start producing pointing gestures mainly in combination with speech, although from 12 months onwards they can combine them for social purposes like information highlighting (Butcher & Goldin-Meadow, 2000; Igualada et al., 2014; Murillo & Belinchón, 2012; see also Chapter 2 in this dissertation). These early pointing-speech combinations already have an adult-like pattern of prominence alignment in that the gesture stroke coincides with the accented syllable (see Chapter 2 in this dissertation).

An unexplored issue is whether younger infants are sensitive to this precise alignment between the prominent part of co-speech gestures (i.e., the stroke) and the prominent parts of speech (i.e., prosodically accented syllables) before they produce their first aligned pointing-speech combinations. That is, are younger infants sensitive to the fact that gesture and speech prominences occur together in speech-accompanied gestures?

Studies on infants' sensitivity to audiovisual (A-V) speech synchrony indicate that 4- to 10-month-old infants can detect an A-V desynchronization (with speech sounds and lip movements out of synchrony) in both isolated syllables and fluent speech

(Lewkowicz, 2010; Pons & Lewkowicz, 2014). On the other hand, but crucial for the current study, infants' perception and detection of prosodic speech prominence have been reported to appear later. It is widely accepted that infants can discriminate basic word stress patterns from birth (Sansavini, Bertoncini, & Giovanelli, 1997). The ability to process or represent word stress patterns has been observed after 6 months of age (Höhle et al., 2009). However, when using more complex or variable stimuli, it is only around 9 months of age that discrimination is observed (Pons & Bosch, 2010; Skoruppa et al., 2009, 2013).

The current study is aimed at exploring whether 9-month-old infants are sensitive to the alignment between gesture prominences and speech (prosodic) prominences in co-speech pointing gestures. We predicted that infants would be sensitive to the alignment between prominences at this early age, before they produce pointing-speech combinations.

3.2. Method

3.2.1. Participants

To test our prediction, we tested twenty-four full-term 9-month-old Catalan-learning infants. The infants had an average age of 9.01 months (range: 256-287 days). Twelve additional infants were tested but were not included in the final sample because of crying or

fussiness (5 infants), failure to habituate (5), and experimental error (2). Participants were recruited at the maternity unit of the Hospital Sant Joan de Déu in Barcelona, Spain. Parental consent was obtained before running the experiment.

3.2.2. Materials

The stimuli consisted of multimedia movies which were constructed with Premiere Pro CS5.5 (Adobe Corporation). The movies were video clips of a woman producing a pointing gesture accompanied by a disyllabic word produced in an infant-directed manner. The woman appeared sideways in the right part of the screen, pointed to the left part of the screen and it covered her mouth with the hand not used for pointing to prevent infants from looking at her lip movements. Eighteen words (half iambs and half trochees) were used. All words were high-frequency, common words for the infants according to the MacArthur-Communicative Development Inventory (MCDI) (Fenson et al., 1994). There were two types of video clips: the aligned clips, in which the gesture stroke (those video frames that capture maximum extension of the arm during the pointing gesture) coincided with the accented syllable of the pointing-accompanying word; and the misaligned clips, in which the gesture stroke coincided with the unaccented syllable of the pointing-accompanying word (see Figure 12). The misaligned clips were created using Premiere Pro CS5.5 by decoupling video and audio tracks and then displacing the video track backwards (in the

case of iambs) or forwards (in the case of trochees) so that the gesture stroke coincided with the unaccented syllable.

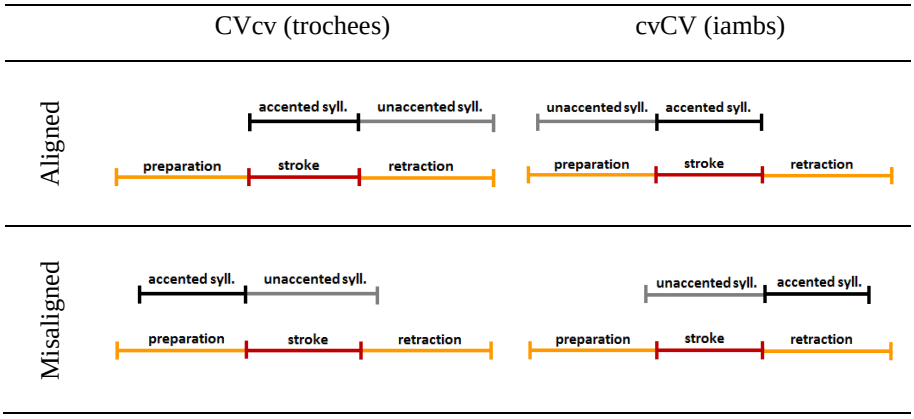

Figure 12. Schematic diagram showing how audio and video tracks used in the experiment were aligned or artificially misaligned: the black and grey lines represent the speech signal, while the red and orange lines the gesture phases associated with it.

3.2.3. Procedure

Infants were tested in a dimly lit and sound-attenuated laboratory room, seated in a high chair facing a LG 50" TV screen at a distance of approximately 130 cm. The experiment was controlled by the experimenter from an adjacent room using Habit 2002 software (Cohen, Atkinson, & Chaput, 2000) running on a Power Mac G5. The infants' looking behavior was video-recorded for subsequent analysis. The habituation/test procedure was used to test for the detection of prosody-gesture alignment. The habituation

phase consisted of the presentation of 15-second trials, each with three aligned video clips. During the habituation phase infants were presented with both iambic and trochaic stimuli (all words presented within a trial had the same stress type). The habituation criterion was set such that infant looking had to decline during a three-trial block to 60% of the total looking time observed during the longest block of three trials. When infants reached this criterion, the habituation phase ended and the test phase began. In the test phase four trials were presented, each consisting of four video clips. Two test trials contained aligned clips (one with iambic words and the other with trochees) and the other two test trials contained misaligned clips (one with iambic words and the other with trochees). These trials were presented in counterbalanced order across infants.

3.3. Results

To determine whether infants were sensitive to the prosody-gesture misalignment we compared the infants' duration of looking time at each test trial. We submitted the data from the four test trials to a 2 × 2 × 4 mixed, repeated-measures ANOVA, with 'stress pattern' and 'alignment' as within-subjects factors and test-trial order as the between-subjects factor. This analysis yielded only a significant main effect for 'alignment' ($F(1, 20) = 7.262$, $p = .014$, partial $\eta^2 = .266$), indicating that infants detected the difference between the

aligned and misaligned stimuli, and that this detection was not affected by the lexical stress pattern of the words (see Figure 13).

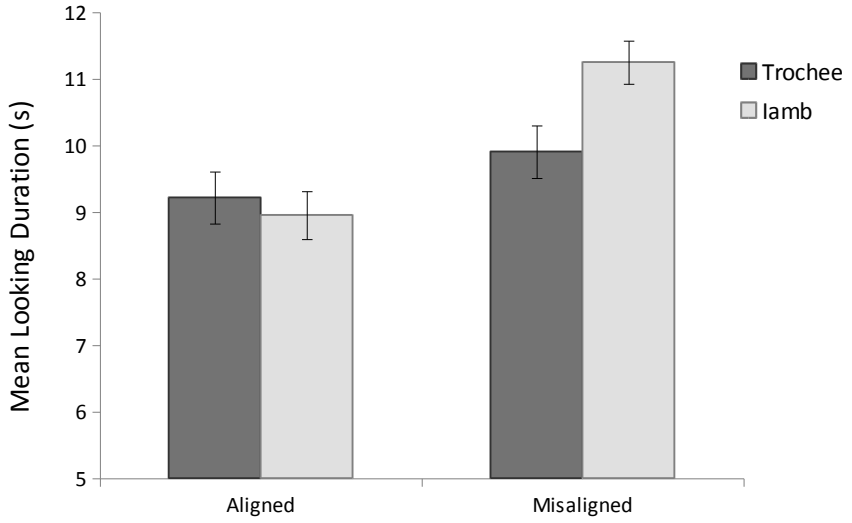

Figure 13. Mean looking times during test trials by 9-month-old Catalan-learning infants. Error bars represent standard error.

3.4. Discussion and conclusions

Two main implications can be derived from our results. First, our results suggest that infants' early sensitivity to prominence is multimodal, because their ability to discriminate between word-initial and word-final prosodic prominence (Pons & Bosch, 2010; Skoruppa et al., 2009, 2013) can also be applied to pointing gestures. At 9 months of age infants not only know that prosodic

prominence occurs at distinct positions within the word but they also know that the prominence in pointing gestures has to coincide with it. Second, our results show that infants are aware of the adult-like timing of the gesture-speech combinations well before they are able to produce these combinations.

Our results contribute to the literature on gesture and speech integration by revealing that gesture and speech combinations are perceptually integrated in infants from very early stages in language and cognitive development. These findings support theoretical accounts which posit that both modalities are part of the same single system in human communication (Kelly et al., 2010; McNeill, 1992). Future research should investigate whether the early sensitivity seen in infants to the prosodic-gesture alignment takes into account linguistic/semantic constraints or is based purely on the perception of systematic alignment patterns. Recent research with adults suggests that in natural discourse the timing of gesture-speech combinations can be influenced by the semantic coordination between the two modalities (Bergmann, et al., 2011; Esteve-Gibert, Pons, Bosch, & Prieto, 2014), that certain gestures such as negation gestures and emotion gestures do not seem to follow the prosodic timing constraints (González-Fuente, Escandell-Vidal, & Prieto, 2014, Harrison, 2010), and also that different languages align gestures differently in speech depending on the semantics of the word involved (Alferink & Gullberg, 2014). If the infants' sensitivity to gesture-speech combinations is already constrained by semantic and linguistic factors we should also

observe an early sensitivity to these effects, albeit perhaps in older infants.

CHAPTER 4: PROSODY SIGNALS THE EMERGENCE OF INTENTIONAL COMMUNICATION

4.1. Introduction

A number of studies have investigated early prosodic patterns in babbling infants. Some of them have focused on the presence or absence of language-specific prosodic patterns in terms of contour direction, metrical bias, or syllable duration (Davis et al., 2000; Engstrand, Williams, & Lacerda, 2003; Kent & Murray, 1982; Levitt & Utman, 1992; Lieberman, 1967; Mampe et al., 2009; among others). Although it is well known that adults use prosody to express communicative intentions, attitudes, and meanings, this first group of studies investigated prosodic development irrespective of the potential differences in the pragmatic meaning of the vocalizations. A second group of studies did not incorporate intentionality as a factor in their analysis of prosodic development but, when discussing results, they stated that the differences they found in contour direction could be due to communicative purposes (Whalen, Levitt, & Wang, 1991) or to a dynamic relationship between physiological constraints and emotional experience (Snow, 2006; Snow & Balog, 2002).

A third group of studies, however, investigated the emergence of communicative intention in relation to prosody. Many of them have analyzed infants at the one-word stage, finding that at this stage

infants produce adult-like prosodic contours to express distinct pragmatic intentions (Astruc et al., 2013; Balog & Brentari, 2008; Balog, Roberts & Snow, 2009; Flax, Lahey, Harris & Boothroyd, 1991; Furrow, 1984; Furrow, Podrouzek & Moore, 1990; Galligan, 1987; Marcos, 1987; Prieto et al., 2012; Vihman et al., 1998; Vihman & DePaolis, 1998). In a longitudinal study from the babbling stage to the one-word and two-word, Halliday (1975) analyzed his son's early pitch contours from 9 months to 2 and a half years of age and discovered that different vocal expressions were able to convey distinct functions. Results showed that the infant 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. Also, he found that at 1 year of age his son produced requests with rising tones. Within this last group of studies investigating prosodic development with respect of intentionality, only a few of them have analyzed infants during the pre-babbling and babbling periods. D'Odorico and Franco (1991), for instance, acoustically analyzed the vocalizations produced by five Italian-learning 4- to 11-month-old infants, in terms of mean F0 values, maximum and minimum pitch, melody type structure 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 infants were 9 months of age. Thus, 4-6 month-old-infants of age used different contour directions when producing a VIM and a VSE; 6-8 month-old infants assimilated categories VSE and VAM; and at 8-12 months of age VIM vocalizations could not be distinguished from the other vocalizations. The authors hypothesized that a infant's ability to acoustically distinguish between categories tends to disappear as age increases. Therefore, the authors concluded that until 9 months of age but not thereafter infants show a selective production hypothesis, 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.

Papaeliou et al. (2002) study represented a step forward in identifying the prosodic cues that infants use in the babbling period to express intentionality. They examined the acoustic patterns of six English-speaking infants from 7 to 11 months of age and they acoustically analyzed vocalizations expressing either emotions or communicative functions. According to Trevarthen (1990), vocalizations expressing emotions identify the quality of communication, whereas vocalizations expressing communicative functions identify the direction and purpose of communication. They analyzed the following features in the vocalizations: duration;

initial, final, peak, lowest, and mean F0 values; range of F0; standard deviation of F0; ratio of standard deviation of F0; and duration of the vocalization. The meaning of the vocalizations was assigned by interviewing mothers about the meaning they would attribute to their infant's vocalizations, a system that, according to the authors, simulates the natural conditions of communication. They found that prosodic patterns were different when vocalizations conveyed communicative functions from when they expressed emotions: vocalizations carrying communicative functions were shorter, with lower F0 values, and had greater intensity than vocalizations expressing emotions. Similarly, Papaeliou and Trevarthen (2006) found evidence that prelinguistic vocalizations can be a tool for both communicating and thinking. They observed four English-speaking 7- to 11-month-old infants and classified their vocalizations as 'communicative' or 'investigative' according to concurrent non-vocal behaviors. They considered a vocalization to be investigative if the infant was holding an object, inspecting an object, or completing a task; they considered it communicative if the infant was interacting with an adult, pointing, directing eye-gaze at the adult, and reaching or giving something. They observed that infants displayed different prosodic patterns when vocalizations were classified as communicative relative to when they were classified as investigative: compared to investigative vocalizations, communicative vocalizations had a higher mean and maximum F0, higher standard deviation of F0, and shorter duration.

All in all, very few studies have investigated infants' use of prosodic contours to express distinct pragmatic functions when

infants are younger than 12 month-old. Even though it has been found that infants can produce adult-like prosodic patterns at the one-word stage, little is known about whether intentional differences influence the prosodic patterns of vocalizations at an earlier age. The purpose of the present study is to investigate whether infants express intentionality by means of prosodic cues when they are still not able to produce words; and, if they do so, how they do it.

Thus, the goal of this study is twofold. First, it seeks to investigate whether babbling infants use prosodic cues such as pitch range or duration to distinguish between intentional and non-intentional vocalizations during the second half of their first year, since it is during this period that infants start communicating intentionally (e.g., Piaget, 1936; Trevarthen, 1977; 1979; 1982). We analyzed a total of 2,701 naturalistic vocalizations recorded from four Catalan-speaking infants at 7, 9, and 11 months of age. Following Papaeliou and Trevarthen (2006), our hypothesis was that infants' non-intentional vocalizations would be produced with a narrower pitch range and longer duration than intentional utterances. If this hypothesis were corroborated, Papaeliou and Trevarthen's (2006) results would be strengthened with a language other than English and with a wider corpus, since that study tested only 193 vocalizations and the current study includes over 2,000 vocalizations. Second, our study aims at discovering whether babbling infants are able to use such prosodic cues (pitch range and duration) consistently in order to express distinct pragmatic functions such as request, discontent, response, or statement. This

second goal represents a step forward in the analysis of how the development of prosody is related to the emergence of communicative intention, given that previous studies on prosodic development of babbling infants did not take into account pragmatic considerations. In general, we hypothesized that (1) babbling infants will display a consistent use of prosodic cues to distinguish intentional from non-intentional vocalizations, based on results found in previous studies, and that (2) when intending to communicate, babbling infants will also select prosodic cues to convey specific pragmatic intentions. Previous studies found that prosody is used by babbling infants to signal the intentional status of a vocalization. Therefore, the corroboration of the first hypothesis would confirm results from prior studies. However, to our knowledge, no studies have investigated whether babbling infants use prosody to distinguish between specific intentions, even though the babbling period in language development is known to coincide with the infants' development of intentionality. Verifying our second hypothesis, then, would suggest that prosody is a tool that infants use during the babbling period to express communicative intentions.

4.2. Method

4.2.1. Participants

Four Catalan-learning infants participated in the study, two male (Bi and Ma) and two female (An and On). Infants were recorded

weekly from 7 to 11 months of age. The present study analyzes infants' vocalizations at ages 7, 9, and 11 months. If we take Piaget's four stages of cognitive development as a reference, the period of interest would be included in the late 3rd and the 4th sub-stages of the sensorimotor stage. It is during these sub-stages that intentionality and logic emerge, starting with intentional grasping of a desired object and differentiating between means and goals, and ending up with the coordination of schemes and intentionality, and planning steps to achieve an objective.

All parents of the four participants speak exclusively Catalan to their infant and to each other. Parents were asked about their linguistic habits through a questionnaire, and results showed that all four mothers have Catalan-speaking parents, have lived in Catalonia all their lives, and have Catalan as their first language (L1). They use Catalan in all dealings with their family, work colleagues, and friends. As for fathers, three of them have Catalan-speaking parents, and have always lived in Catalonia. Catalan is their L1 as well as the vehicular language for family, work, and friends. An's father, however, has Spanish-speaking parents and uses Spanish as the primary language for communicating with his parents and work colleagues. However, he speaks and writes Catalan fluently, and uses it with his wife, daughter, and friends. The infants come from four small towns in the same region of Catalonia, Alt Penedès, located 50 km to the south of Barcelona. According to the information available from the official statistics website of Catalonia (www.idescat.cat, Linguistic census from 2011), in three of these towns Catalan is spoken regularly by about 90 percent of

the population, and in the fourth town Catalan is spoken by 80 percent of the population. Thus, it may be safely assumed (and also according to the parents' reports) that there is very little Spanish influence in the infants' linguistic input, since infants are not exposed to Spanish at home and hear very little of it outside the home.

4.2.2. Data collection

All infants were video-recorded at their homes during weekly 30-minute sessions between 7 and 11 months of age using a SONY camera, model DCR-DVD202E PAL. Thus, they were all recorded three to five times per month, except for Bi at 9 months and On at 11 months, who were recorded only twice during those months due to illness. Recordings were made by the first author of this study, who was previously acquainted with the families and infants. Infants were always recorded in the same room of their respective homes, typically their living-rooms, during free-play sessions. All infants were recorded as they interacted with their mothers, except for one infant, 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 infant as possible and positioned so that the camera was pointing toward the infant's face.

In order to monitor vocabulary acquisition, the same set of toys was given to the infant in all sessions. The first toy offered, a pyramid of four colored plastic stackable disks with animal heads, was common

to all four infant subjects and available to them only during the recording sessions. When subjects lost interest in this toy (which tended to happen after about ten minutes), their parents offered them another toy from the infant's own collection, usually the same toys from one recording session to the next.

From all the weekly sessions recorded during this six-month period, we selected for analysis vocalizations produced when the infants were 7, 9, and 11 months of age. 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 (e.g., Piaget, 1936; Trevarthen, 1977; 1979; 1982).

4.2.3. Data analysis

The approximately 18 hours of recordings were segmented into 2,946 vocalizations. From these, 245 were excluded from the analysis because of the following circumstances: (1) infant and parent overlapped when vocalizing, (2) ambient noise was too loud, (3) the infant vocalized while having an object inside his/her mouth, or (4) the sound did not show a visible trace on the spectrogram. This yielded a corpus of 2,701 vocalizations.

Before segmenting the data, we established the unit of analysis of our study. Following Papaeliou and Trevarthen (2006), two

utterances were considered distinct vocalizations if they were separated by 50 ms or more. Additionally, when there were more than 50 ms between two vocalizations, but their prosodic contours were linked by a sustained fall at the end of the first vocalization followed by a second vocalization starting at that sustained F0 level, they were not separated but considered the same vocalization.

a) Pragmatic analysis

All vocalizations were first annotated by one coder in terms of the communicative function they conveyed using the Phon software system (Rose et al., 2006). Different authors have dealt with the classification of pragmatic functions of early vocalizations in different ways. As noted above, D'Odorico and Franco (1991) used the terms 'vocalizations during infant manipulation of a toy', 'vocalizations during shared experience' (manipulating a toy but looking at the adult), 'vocalizations during adult manipulation of a toy', and 'vocalizations during exchanges with the adult' (neither of them is manipulating the toy but they are both looking at each other). Blake and Boysson-Bardies (1992) classified their subjects' vocalizations using the following labels: fine object manipulation, gross object manipulation, upright movement, confined movement, request, comment, book-reading, demonstrative, response to adult's utterance, give and take, rejection-protest, or physical interaction. In addition, Sarriá (1991) and Karousou (2003) used these categories: request (object, help, or attention), rejection, protest, satisfaction,

question (what, where, and how), statement, proto-conversation, narration, interactive game, imitation, non-social, or greeting.

Since the first aim of our study was to discover whether the vocalizations of Catalan-babbling infants conveying communicative information are different from vocalizations that did not intend to communicate information, we first classified our data into one or the other, labeled respectively 'intentional' or 'non-intentional'. Following Papaeliou and Trevarthen (2006), a vocalization was considered to be non-intentional if the infant was holding an object, inspecting an object, or completing a task; a vocalization was considered to be intentional if the infant was interacting with an adult, pointing, directing eye-gaze at the adult, and reaching or giving something. Thus, the distinction between intentional and non-intentional vocalizations relied mostly on gestural cues, as well as context and parental reactions before or after the vocalization.

Apart from the labels 'non-intentional' and 'intentional', an extra category was used to classify all those utterances that were difficult to label. Thus, 'not clear' was the label used when visual cues were not clear enough to decide whether a vocalization was intentional or not. For instance, when the infant was vocalizing but her hand or face was not visible in the video (e.g. behind the sofa), it was included in the 'not clear' group. The presence of this third category enhances the reliability of the results, since no vocalization was forced to fit into one of the other two categories described above. A total of 324 vocalizations were labeled as 'not clear' following this criterion. Thus, of a sum of 2,701 recorded vocalizations, our

analysis yielded a total of 1,676 intentional vocalizations, 701 non-intentional vocalizations, and 324 vocalizations whose purpose was ‘not clear’.

In order to test the second hypothesis, i.e. whether infants select certain prosodic cues to express distinct pragmatic functions, all intentional vocalizations were further classified into narrower categories depending on the specific pragmatic functions the infant was judged to be performing. The pragmatic functions adopted were based on Sarriá (1991) and Karousou (2003). The specific intentions used were discontent (the infant expressed ‘sadness’ actively), request (the infant wanted the other person to do something), response (the infant reacted to a stimulus, either a verbal stimulus uttered by an adult or an action performed by the adult), satisfaction (the infant expressed happiness about the current situation), statement (the infant vocalized simply because (s)he wanted the adult to know something), surprise (the infant wished to express the idea that an unusual or unexpected event had occurred), and vocative calling (the infant called somebody). Hence, the pragmatic analysis consisted not only of deciding whether a vocalization was intentional or non-intentional but also of deciding whether that vocalization bore a specific intentionality. In order to screen out the potential influence of prosodic cues in the audio material, this specific classification was performed only when the recording displayed clear contextual and non-vocal information. All those intentional vocalizations that were impossible to classify further into one specific pragmatic meaning were included in a category called ‘fuzzy intention’. Thus, when a vocalization was

clearly intentional but too fuzzy to fit in any of these specific pragmatic categories, it was labeled as ‘fuzzy intention’. Such cases represented 745 out of the 1,676 intentional vocalizations. In sum, all vocalizations relevant for our study were first classified as ‘intentional, ‘non-intentional’, or ‘not clear’. Next, the group of ‘intentional’ vocalizations was further subdivided into the specific pragmatic functions. These classifications were conducted on the basis of audio and visual cues in the recordings. Importantly, in order to minimize the potential influence of prosodic/acoustic cues in determining the intentional status and specific intention of vocalizations, the pragmatic and gestural analyses of all vocalizations (performed independently using Phon) were performed prior to the acoustic analysis (performed independently using Praat) (see the following sections).

To test the reliability of the pragmatic coding, an inter-transcriber reliability test was conducted with a subset of 20% of the total number of vocalizations in the target materials (which represented a total of 540 utterances), making sure that all infants and ages were uniformly represented. Three independent coders labeled a random selection of 20% of the data in terms of intentionality and specific pragmatic intentions. The overall agreement was 82% when deciding whether the vocalization was intentional or not, and 74% when deciding on specific pragmatic intentions. The fact that the overall agreement was lower when rating specific pragmatic intention than when rating the intentional status might be due to the fact that in the former case raters had to choose among a considerably higher number of categories or because some of the

specific intentions were more difficult to categorize. For instance, raters sometimes found it difficult to distinguish between the categories ‘discontent’ and ‘request’ because in some cases a infant might urge the adult to do something while expressing sadness. All in all, we think that these scores reveal a substantial agreement among raters and are comparable with other studies’ scores (Chen & Kent, 2009; Papaeliou & Trevarthen, 2006). Chen and Kent (2009), for instance, achieved an overall agreement of 84% in their inter-transcriber reliability test.

b) Gesture analysis

The gestural analysis was performed in parallel with the pragmatic analysis described above. As is well known, infants begin to gesture very soon in order to influence the mental state of others, i.e. because they want others to do, know, or feel something (Tomasello et al., 2007). The first communicative gestures that typically developing infants produce are deictics such as pointing, giving, showing, or requesting (Bates et al., 1979; Iverson & Goldin-Meadow, 2005; Ozçalişkan & Goldin-Meadow, 2005; Sansavini, Guarini & Stefanini, 2010; Tomasello et al., 2007). Each vocalization was annotated in terms of the gestures displayed by infants when vocalizing, using the Phon software system (Rose et al., 2006). All vocalizations were labeled with gestural information regarding gaze direction, manual gestures, and facial gestures. A simplified version of Allwood, Cerrato, Jokinen, Navarretta and

Paggio's (2007) categories was adopted in the present study for the annotation of infants' gestures: hand gestures were defined in terms of handedness (single hand, both hands), hand trajectory (up, down, sideways, etc.), and their semiotic and communicative value; facial gestures were defined in terms of general face, position of the eyebrows, eye position, gaze direction, form of the mouth, head position, and their semiotic and communicative value. This codification system was chosen because it enabled us to code gestures independently of their possible meaning or function, using the system's labels regarding the form of the gesture. Table 8 shows the gesture categories used in our study.

Gaze direction	absent gaze
	gaze at camera
	gaze at object
	gaze at parent
Manual gestures	clapping hands
	extending arms
	hugging parent
	manipulating object
	moving arms
	pointing at object
	moving hands
	shaking arms
	no specific manual gesture
Facial gestures	furrowing brows
	opening eyes
	closing eyes
	opening mouth
	closing mouth
	pouting
	shaking head

	smiling
	rising eyebrows
	no specific facial gesture

Table 8. Gesture categories used in the gesture analysis.

c) Acoustic analysis

The main aim of this study was 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 WAV format) from our Phon corpus and analyzed them with the Praat software package (Boersma & Weenink, 2012). Also, no information on infant, age, pragmatic intention, or gesture was at the coder's disposal when annotating the acoustic measures, in order to guarantee that there would be no influence of pragmatic coding on the determination of acoustic parameters.

Two prosodic features were manually labeled: duration and pitch range, i.e. start and end points of vocalizations, and pitch maximum and minimum points. The aim was to analyze the 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). As for pitch range, an overview of the data indicated that the best way to obtain this measure was to select three pitch points from the fundamental frequency contour. These three pitch points were distributed along the fundamental

frequency line and included the lowest (F0 min) and highest points (F0 max). The first pitch point (p1) was selected at the onset of vocalization, since this point is usually referred to as the reference level of the speaker; the second pitch point (p2) was generally selected at a point in the middle of the F0 contour; and finally, the third point (p3) was usually selected at the end of the vocalization. However, when the lowest or highest pitch values did not appear at the very beginning, at the very end, or right in the middle of the vocalization, the points selected were moved according to our needs in order to make them coincide with the lowest and highest point.

In percentages, the lowest F0 point was mostly located at p3 (50.47% of cases) or p1 (40.02% of cases); the lowest F0 point was located at p2 for just 9.51% of the vocalizations. The highest F0 point was located at p2 in 72.75% of cases, and was less frequently located at p1 (16.66%) or p3 (10.59%). When these points were annotated, the pitch maximum and pitch minimum values were extracted using a Praat script, and the pitch range was calculated by subtracting the pitch minimum from the pitch maximum. In order to compare different pitch ranges across the four infants, pitch values were extracted in semitones rather than in Hz.

Additional considerations for determining the F0 index measurements were as follows: (a) when the vocalization had more than one peak point at the same level, the last point was selected; (b) if the vocalization displayed no clear peak, a pitch point in the middle of the vocalization was selected.

To obtain the total duration of the vocalization, the first point (t1) and last point (t2) in the F0 line of the vocalization were selected. Two sounds were considered to be distinct vocalizations if they were separated by at least 50 ms (Papaeliou & Trevarthen, 2006). When there was 50 ms between two vocalizations but they were prosodically linked, they were considered one vocalization. Figure 14 illustrates the annotation of pitch range and duration. The first tier was used to annotate start and end time of the vocalization (t1, t2), and the second tier was used to annotate the three index pitch points (p1, p2, p3) to later calculate pitch range values. The upper graph is an example of a non-intentional vocalization and the lower graph is an intentional vocalization.

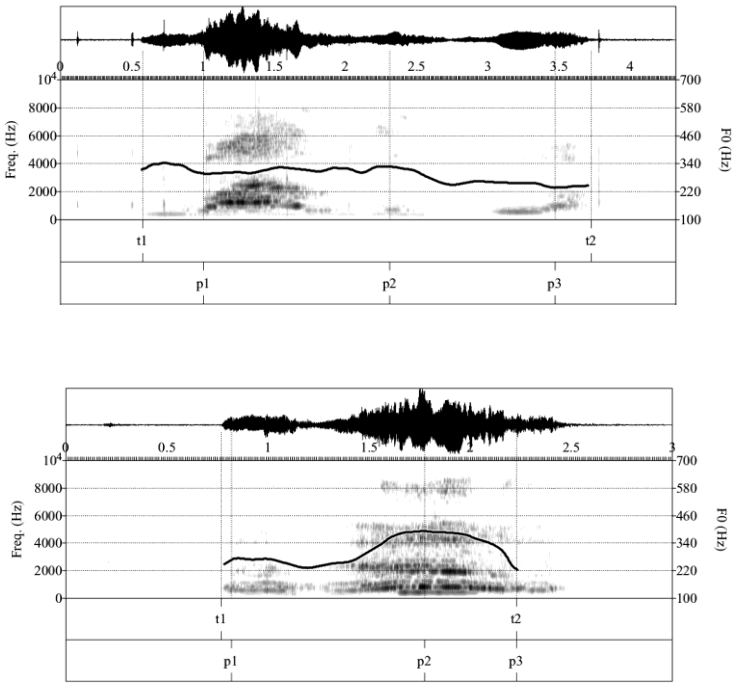

Figure 14. Example of an annotated non-intentional vocalization (top) and an intentional vocalization (bottom) performed by Ma at 9 months of age.

4.3. Results

This section includes two different parts. The first part presents the results of the analysis of the potential effects of the intentionality on prosodic cues (i.e. pitch range and duration). The second part presents the results regarding the potential effects of the pragmatic function on prosodic cues (i.e. pitch range and duration). All statistical analyses in this article were performed by applying a linear mixed model (LMM; West et al., 2007) using SPSS Statistics 15.0 (SPSS Inc., Chicago IL). West et al. (2007) state that LMMs are the appropriate model for analyzing unbalanced longitudinal data, since they allow for subjects with missing time points (i.e. unequal measurements over time for individuals), have the capacity to include all observations available or all individuals in the analysis, and cope with missing data at random. As West et al. (2007) point out, linear mixed models can accommodate all of the data that are available for a given subject, without dropping any of the data collected from that subject.

4.3.1. Prosodic cues and intentionality

Table 9 and Figure 15 show a general overview of the data included in the analysis. Table 9 displays the number of vocalizations produced by each infant at each age, and their classification according to the intentional status. Figure 15 shows the percentage of 'intentional, 'non-intentional', or 'not clear' vocalizations across the different ages. These results reveal that infants produce more

intentional vocalizations than non-intentional vocalizations at all ages and that such expressions increased longitudinally: at 7 and 9 months of age intentional vocalizations approximately double the number of the non-intentional ones, and at 11 months of age the intentional vocalizations are four times more frequent than the non-intentional ones. They also show that 12% of the total number of vocalizations could not be identified as being either intentional or non-intentional.

Chi-squared tests of independence were carried out in order to investigate whether the proportion of 'intentional and 'non-intentional' vocalizations differed from each other and across ages. Results showed that the proportion of intentional and non-intentional vocalizations was statistically different at all ages ($\chi^2(1, N = 610) = 41.512$, $p < .001$ at 7 months of age, $\chi^2(1, N = 726) = 57.322$, $p < .001$ at 9 months of age, and $\chi^2(1, N = 1041) = 35.308$, $p < .001$ at 11 months of age). As for the potential significant difference among proportions of intentional and non-intentional vocalizations across ages, the chi-squared tests revealed that the proportions of intentional vocalizations differed significantly at all ages: from 7 to 9 months of age ($\chi^2(1, N = 859) = 7.728$, $p = .005$), from 7 to 11 months of age ($\chi^2(1, N = 1211) = 160.861$, $p < .001$), and from 9 to 11 months of age ($\chi^2(1, N = 1292) = 100.465$, $p < .001$). In contrast, the proportion of non-intentional vocalizations varied significantly only from 9 to 11 months of age ($\chi^2(1, N = 475) = 4.651$, $p = .031$), and not from 7 to 9 months ($\chi^2(1, N = 487) = 2.667$, $p = .102$), nor from 7 to 11 months ($\chi^2(1, N = 440) = 0.276$, $p = .600$).

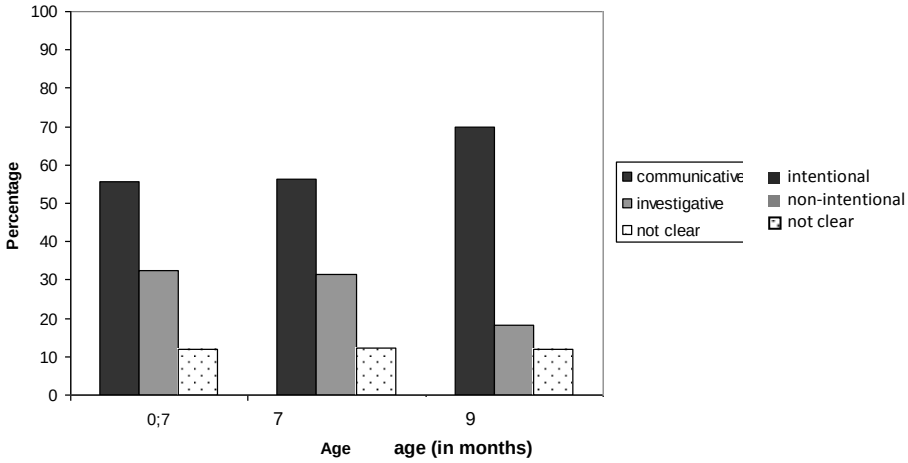

Figure 15. Percentages of 'intentional, 'non-intentional', and 'not clear' vocalizations across the different age groups.

	7 months	*9 months*	*11 months*	*Total*
intentional vocalizations	384	465	827	1,676
non-intentional vocalizations	226	261	214	701
'not clear' vocalizations	85	102	137	324
TOTAL	695	828	1178	2,701

Table 9. Number of vocalizations classified in terms of intentional status and age.

In the following sections, we discuss the effect of the intentional status on pitch range and then we move on to its effects on duration. All statistical analyses were performed excluding outliers (N = 13) and vocalizations labeled as 'not clear' (N = 324).

a) Pitch range and the intentionality

The relationship between pitch range and the intentional status of vocalizations was analyzed using linear mixed model analysis (LMM). Pitch range (in semitones) was the dependent variable, and fixed factors were age (3 levels: 7, 9, and 11 months), intentionality (2 levels: intentional and non-intentional), and the interaction between age and intentionality. Infant was classified as a random factor and not a fixed factor because the purpose of the study was not to investigate individual differences and also because previous analyses of the data revealed that the variable 'infant' did not have a significant effect on the results. The analysis revealed a statistically significant effect of intentionality on the pitch range ($F(1,2073) = 12.690$, $p < .001$). No significant effects of age were found on pitch range ($F(2,2047) = 0.816$, $p = .442$), and results on the interaction between intentionality and age were also non-significant ($F(2,2073) = 0.214$, $p = .807$). Figure 16 shows the pitch range displayed by intentional and non-intentional vocalizations at the three ages analyzed.

b) Duration and intentionality

The relationship between duration and the intentional status of the vocalization was tested using LMM analysis with duration (in milliseconds) as the dependent variable, and age (3 levels: 7, 9, and

11 months), intentional status (2 levels: intentional and non-intentional), and the interaction between age and intentional status as fixed factors. Again, infant was classified as a random factor and not a fixed factor for the reasons stated above. The statistical analysis showed that duration was significantly affected by age ($F(2,2072) = 22.602$, $p < .001$) as well as the intentional status of the vocalization ($F(1,2072) = 57.732$, $p < .001$). The interaction between age and intentionality, however, was not significant ($F(2,2072) = 0.879$, $p = .415$).

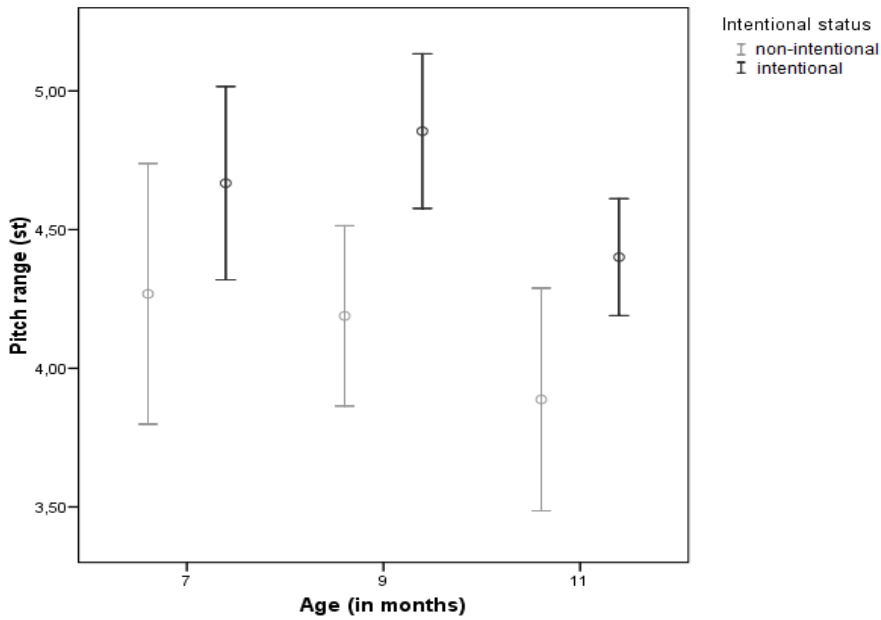

Figure 16. Error bars of the pitch range of vocalizations (in semitones) as a function of intentional status and infants' age.

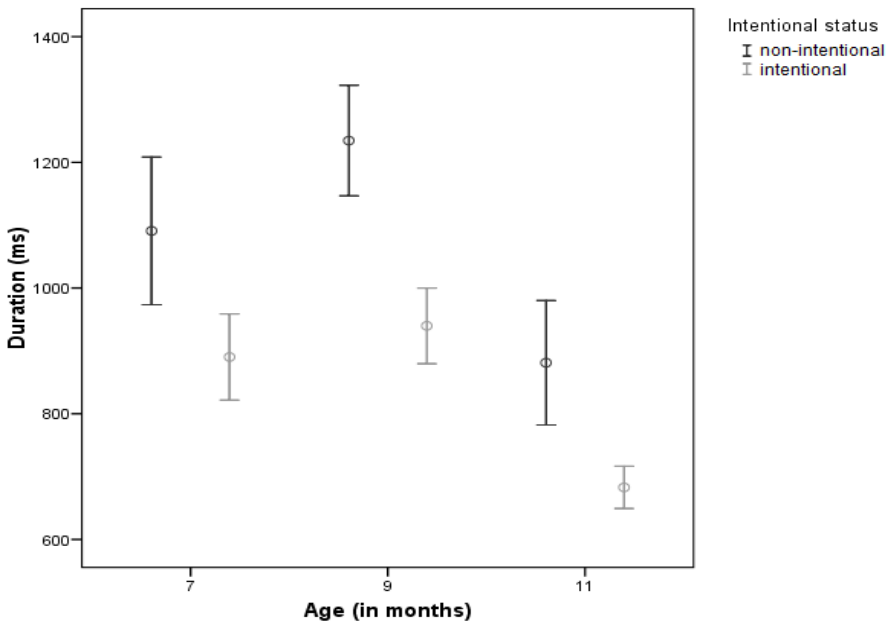

Figure 17. Error bars of the duration of vocalizations (in milliseconds) as a function of intentional status and infants' age.

Bonferroni-corrected pairwise comparisons revealed that the mean duration differed significantly from 7 to 11 months of age ($p < .001$) and from 9 to 11 months of age ($p < .001$) but not from 7 to 9 months of age ($p = .062$). Thus, results for duration in relation to the intentional status of the vocalizations were more robust at 9 and 11 months of age than at 7 months of age. Figure 17 displays the error bars of the total duration of vocalizations (in milliseconds) as a function of intentionality. These results show that at all ages intentional vocalizations tended to be shorter than non-intentional vocalizations. It can also be observed that this difference is more prominent for some ages than others: at 7 months of age the mean

duration of an intentional vocalization is 890.30 ms (*SD* = 631.668) compared to 1090.94 ms (*SD* = 770.798) for a non-intentional vocalization; at 9 months of age the mean duration of an intentional vocalization is 939.83 ms (*SD* = 626.628), compared to 1234.57 ms (*SD* = 655.421) for a non-intentional vocalization; and at 11 months of age, an intentional vocalization lasts a mean of 682.90 ms (*SD* = 474.791) compared to 881.16 ms (*SD* = 679.777) for a non-intentional vocalization.

In sum, statistical analyses of the data showed that pitch range and duration were both significantly affected by the intentional status of the vocalization. As for pitch range, vocalizations displayed a wider pitch range when infants were intentional than when they were performing non-intentional vocalizations. In terms of duration, intentional vocalizations were shorter in general than non-intentional ones. Yet our results also seem to show that the duration cue was not controlled until infants were 9 months of age. To clarify the picture, in the next section we will investigate whether the specific pragmatic meaning conveyed by the intentional vocalizations has an effect on the pitch range and duration patterns.

4.3.2. Prosodic cues and specific pragmatic intentions

We investigated the prosodic cues within the intentional vocalization group by investigating how pitch range and duration patterns of the vocalization were influenced by the specific

pragmatic function displayed. Table 10 shows the number of vocalizations analyzed classified in terms of age and specific intentional purpose. As the table shows, vocalizations expressing discontent and satisfaction are the most frequent in the corpus (N = 400 and 191, respectively), followed by statements (N = 143), requests (N = 97), and responses (N = 78). Interestingly, statements, responses, and requests are found more often in the corpus when infants are 11 months of age but not when they are younger, whereas expressions of discontent and satisfaction are regularly produced at the earliest stages analyzed. The fact that 7-month-old infants in our study expressed mainly discontent and satisfaction and that most of the pragmatic intentions did not appear until 11 months of age is similar to what Snow and Balog (2002) and Snow (2006) found in their studies, namely that in 8-month-old infants intonation is still influenced by emotional factors.

Specific intentions like ‘surprise’ and ‘vocative’ were seldom produced in comparison with other pragmatic functions like ‘discontent’ or ‘satisfaction’. The low frequency of occurrence of these two categories (see Table 10) meant that they could not be reliably compared with the other relatively abundant pragmatic functions and we therefore decided to exclude them from further analysis. The table also shows that the group including most vocalizations is the group labeled as ‘fuzzy intention’: the proportion of intentional vocalizations which did not have a clear intention was 51.69% at 7 months of age, 47.95% at 9 months of age, and 39.13% at 11 months of age. As noted above, this group

included all those intentional vocalizations that could not be unambiguously identified as any specific pragmatic function.

		7 months	*9 months*	*11 months*	*TOTAL*
Intentional vocalizations	*discontent*	97	137	166	400
	request	24	30	43	97
	satisfaction	56	45	90	191
	response	4	17	57	78
	statement	5	9	129	143
	surprise	-	3	10	13
	vocative	-	-	9	9
	fuzzy intention	199	222	324	745
TOTAL		385	463	828	1,676

Table 10. Number of vocalizations classified in terms of pragmatic intention and age.

a) Pitch range and pragmatic intentions

The relationship between pitch range and specific pragmatic intention displayed for intentional vocalizations was tested using LMM analysis, with pitch range (in semitones) as the dependent variable, and age (3 levels: 7, 9, and 11 months of age), pragmatic intention (5 levels: discontent, request, satisfaction, response, and statement), and the interaction between age and pragmatic intention as fixed factors. Again, infant was classified as a random factor.

Results revealed a significant effect of specific pragmatic intention on pitch range ($F(4,763) = 4.539$, $p = .001$). No effect of age was found for pitch range ($F(2,729) = 1.544$, $p = .214$), and there was no interaction of age or intention with pitch range ($F(8,784) = 1.356$, $p = .212$).

As Table 11 shows, Bonferroni-corrected pairwise comparisons revealed that there were no significant differences in pitch range across pragmatic intentions, except for expressions of discontent, which vary significantly from expressions of satisfaction ($p = 0.006$). When looking at mean pitch range values with all ages combined, distinct tendencies can be observed across pragmatic intentions: the mean pitch range for expressions of discontent was 5.37 st ($SD = 3.18$), 5.10 st ($SD = 2.85$) for requests, 4.46 st ($SD = 3.09$) for expressions of satisfaction, 3.82 st ($SD = 2.84$) for statements, and 3.73 st ($SD = 2.49$) for responses. Figure 18 shows the different tendencies across pragmatic intentions: expressions of discontent display wider pitch range, requests show a pitch range that is narrower than that of expressions of discontent but wider than that of the other intentions; expressions of satisfaction show a pitch range that is narrower than that of expressions of discontent and requests but wider than that of responses and statements; statements show a pitch range that is narrower than that of expressions of satisfaction but slightly wider than that of responses, and responses are the pragmatic intention that display the narrowest pitch range. Although the differences in mean pitch range are not statistically significant for the most part, they show clear tendencies across pragmatic intentions.

		Pitch range	*Duration*
Discontent	*Request*	1.000	.000**
	Satisfaction	.006*	.000**
	Response	.258	.000**
	Statement	.144	.000**
Request	*Satisfaction*	1.000	.002*
	Response	1.000	.041*
	Statement	.969	.000**
Satisfaction	*Response*	1.000	1.000
	Statement	1.000	.517
Response	*Statement*	1.000	1.000

Note. $*p < .01$, $**p < .001$

Table 11. Statistical *p* values of the pair-wise comparisons of pitch range and duration between pragmatic intentions.

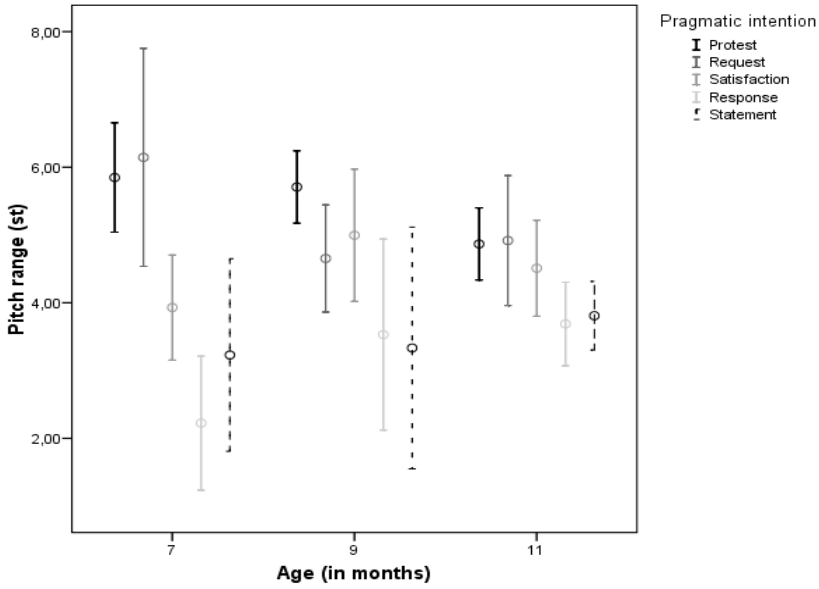

Figure 18. Error bars of the pitch range of vocalizations (in semitones) as a function of the specific pragmatic intention and infants' age.

b) Duration and pragmatic intentions

The relation between duration and specific pragmatic intention displayed in the intentional vocalization was tested once more using LMM analysis. The dependent variable was total duration (in milliseconds), and the fixed factors were age (3 levels: 7, 9, and 11 months of age), pragmatic intention (5 levels: discontent, request, satisfaction, response, and statement), and the interaction between age and pragmatic intention. Infant was once again classified as a random factor. The results showed a significant effect of pragmatic intention on duration ($F(4,787) = 60.841$, $p < .001$). Neither age ($F(2,786) = 1.672$, $p = .189$) nor the interaction between age and intention ($F(8,787) = 1.015$, $p = .423$) had any significant effect on duration.

As Table 11 shows, Bonferroni-corrected pairwise comparisons revealed that some pragmatic intentions varied significantly from each other in terms of duration: vocalizations that express discontent or function as requests were significantly different compared to all other intentions; vocalizations expressing satisfaction had similar duration to responses and statements but differed from expressions of discontent or requests; and responses and statements differed from expressions of discontent and requests. Mean duration values across pragmatic intentions with all ages combined patterned in a similar way to the mean pitch range results reported in the previous section: expressions of discontent showed the longest duration (1241.83 ms, *SD* = 611.02), followed by requests (899.91 ms, *SD* = 513.23); expressions of satisfaction had a

mean duration of 639.71 ms (*SD* = 429.16), while statements lasted 479.59 ms (*SD* = 327.95) on average. The pragmatic intention with the shortest duration (450.40 ms, *SD* = 276.89) was responses. Figure 19 shows these tendencies with error bars. Note that results for the duration of responses and statements at 7 months of age must be treated carefully, since only four vocalizations were classified as responses and only five as statements for that age.

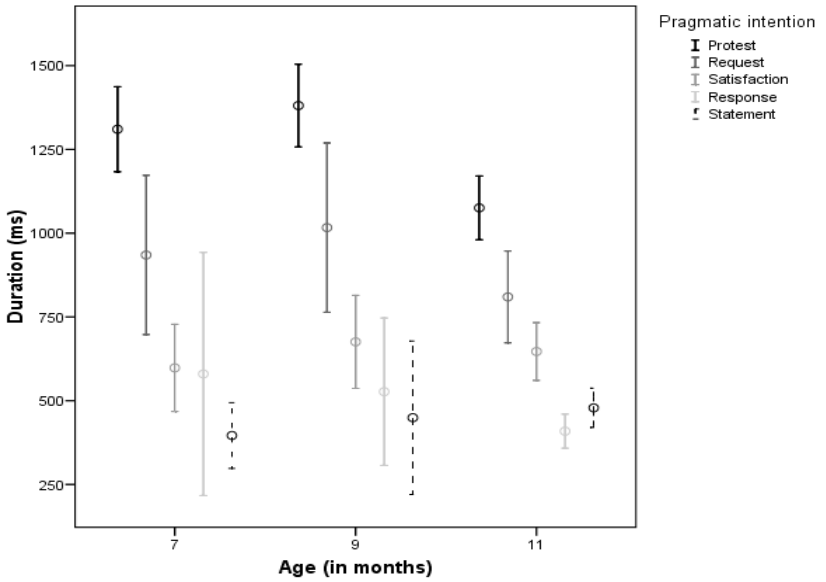

Figure 19. Error bars of the duration of vocalizations (in milliseconds) as a function of the specific pragmatic intention and infants' age.

Hence, the analysis of prosodic cues like pitch range and duration of early vocalizations showed that babbling infants seem to control pitch range and duration as early as 7 months of age. In terms of pitch range, we observed that intentional vocalizations had a wider

pitch range than investigative ones. Further analyses of intentional vocalizations revealed that depending on the pragmatic intention expressed, pitch range tended to be wider or narrower. Thus, expressions of discontent showed significantly wider pitch range than expressions of satisfaction. Further observation of mean pitch range values revealed that although it was not statistically significant, expressions of discontent and requests had wider pitch ranges than responses and statements.

In terms of the duration of vocalizations, our results showed that it was also strongly affected by their intentional status. Non-intentional utterances were significantly longer than intentional ones. Our subsequent analysis of intentional vocalizations, whereby they were categorized into specific pragmatic intentions, showed that the patterns for the duration of the vocalizations were strongly influenced by their specific pragmatic function. Specifically, the shortest vocalizations were responses; statements were slightly longer than responses but still shorter than the other intentions; expressions of satisfaction were longer than responses and statements, but shorter than requests or expressions of discontent. Requests were longer than all the other intentions except for expressions of discontent, which were the longest.

4.4. Discussion and conclusions

This study had two aims: first, to investigate whether infants use specific prosodic cues when attempting to be communicative with

their parents; and second, to investigate whether these babbling infants are able to express specific pragmatic intentions by means of prosodic cues. The longitudinal analysis has revealed that between 9 and 11 months of age infants significantly increase their total number of intentional vocalizations. At 7 and 9 months of age intentional vocalizations are double the number of non-intentional ones; however, at 11 months of age intentional vocalizations are four times more frequent than non-intentional ones. These results support previous studies stating that infants develop intentional communication around 8-9 months of age (Bates et al., 1975; Piaget, 1936; Tomasello, 1993; Vygotsky, 1962).

With respect to our first goal, the prosodic analysis of the data revealed very consistent effects of the intentional status of the vocalizations on prosodic cues such as pitch range and duration. In terms of duration, intentional vocalizations are shorter than non-intentional ones. Even though this tendency was observed at the three ages recorded (namely at 7, 9, and 11 months of age), it was only statistically significant when infants were 9 and 11 months of age. These results suggest that some 7-month-old infants still do not control the use of duration as a prosodic cue to convey intentionality, so it is not until infants are 9 months of age that this ability seems to be acquired. An analysis of a larger database is required to confirm the results on the interaction between duration and intentionality at 7 months of age. As for pitch range, our data has shown that infants produce vocalizations with a wider pitch range when seeking to communicate with their parents and vocalizations with a narrower pitch range when performing

investigative vocalizations. Seven-month-old infants thus seem able to control their vocalizations' pitch range, displaying a wider pitch range when they attempt to communicate and a narrower pitch range when they do not. The patterns of results on pitch range and duration thus replicate Papaeliou and Trevarthen's (2006) conclusions that intentional vocalizations uttered by English-babbling infants tend to have a wider pitch range and shorter duration than non-intentional vocalizations.

Our second goal was to test whether babbling infants were able to use prosodic cues selectively in order to express distinct pragmatic functions well before they produce their first words. First, our data confirm that before producing words, infants are able to communicate intentionally. At 7 and 9 months of age infants are able to communicate with their parents through expressions of discontent and satisfaction, and requests. As their communication skills develop, i.e. at 11 months of age they intentionally produce a wide variety of pragmatic meanings such as expressions of discontent and satisfaction, requests, responses, and statements, apart from random instances of vocatives and vocalizations expressing surprise. These results are consistent with Bates et al. (1975), who based on the Speech Act Theory (Austin, 1962; Bruner, 1975; Searle, 1976), state that before 10 months of age, infants communicate through perlocutions, i.e. communicative acts which have an effect on their listener, but which are not designed as conventions recognized by both speaker and listener; after 10 months of age infants move on to the illocutionary stage, when the infant intentionally uses nonverbal signals to convey requests and to

direct adult attention to objects and events. The fact that 7-month-old infants in our study expressed mainly discontent and satisfaction and that most of the pragmatic intentions did not appear until 11 months of age confirms Snow and Balog (2002) and Snow (2006). These authors found that until around 8 months of age intonation is still influenced by emotional factors.

The results of the acoustic analysis revealed a consistent effect of the pragmatic intention of vocalizations on pitch range and duration patterns. Results of the statistical analyses revealed that utterances classified as discontent had significantly higher pitch range than expression of satisfaction. The observation of the mean pitch range values showed that utterances classified as expressions of discontent and requests have a wider pitch range and longer duration than utterances classified as responses and statements, which are shorter and have a narrower pitch range. Also, expressions of satisfaction lie in the middle ground, as they are shorter than requests and expressions of discontent but longer than responses and statements, and they have a narrower pitch range than expressions of discontent and requests but a wider one than responses or statements. Hence, before the first words are produced, infants are able to select specific prosodic cues to express intentionality in their vocalizations. When infants express discontent or make a request, they consistently use prosodic features like expanded pitch range and longer duration; when they express satisfaction, they use wide pitch range but short duration; and when they produce responses or statements, they use narrow pitch range and short duration.

In sum, our study supports previous research on the prosodic features of prelinguistic vocalizations (D'Odorico & Franco, 1991; Papaeliou et al., 2002; Papaeliou & Trevathen, 2006; Sachs, 1993) in the sense that infants select particular prosodic cues to express intentionality. Our results corroborate the claim that prelinguistic infants produce longer vocalizations with a narrow pitch range when they are playing alone or with a toy and do not interact with their parents. In contrast, their utterances are shorter and show a wider pitch range when interacting with their parents. Yet our results go a step further and show that important prosodic differences are obtained when early vocalizations are related to intentional communication and specific pragmatic intentions. These results thus demonstrate the usefulness of investigating the development of early prosodic patterns at the babbling stage in relation to the development of intentional meaning.

We argued on the basis of our data that before infants produce their first words, they are able to systematically use prosodic cues to express a set of distinct pragmatic meanings. Thus, 9- and 11-month-old infants are able to distinguish expressions of discontent and requests from responses and statements by means of prosody. Recent findings also report the use of adult-like intonational contours to convey specific pragmatic functions in the one-word period (Frota & Vigário, 2008, for Portuguese; Marcos, 1987, for French; Prieto et al., 2012, for Catalan and Spanish). Prieto et al. (2012), for instance, investigated the development of prosodic patterns in four Catalan infants and two Spanish infants and demonstrated that infants at 13 and 15 months of age are able to

produce a set of adult-like intonation contours. Marcos (1987) analyzed the communicative functions of pitch range and pitch direction in 14- to 22-month-old French-learning infants, comparing the prosodic patterns of ten infants when requesting, giving, showing, and labeling. In terms of pitch range, the highest pitch range was found in repeated requests, a somewhat lower range for initial requests, a still lower range for giving and showing, and the lowest range for labeling. For pitch direction, patterns were only clear with requests and labeling, since infants used rising tones when requesting and falling tones when labeling.

Although our babbling data revealed a consistent use of target prosody by young infants, further research is needed to investigate the development of prosodic patterns from the early babbling period to the first-word period by taking into account the communicative uses of language, since it is during the babbling period that infants start using language for communicative purposes. It might well be that the first signs of developmental language impairment can be discernible in the early prosodic patterns that an infant uses when babbling.

CHAPTER 5: PROSODY AND GESTURE HELP INFANTS TO INTERPRET SOCIAL INTENTIONS

5.1. Introduction

Flexibility is a hallmark of human communication. Because one and the same utterance can mean very different things, one needs to infer the underlying communicative intentions of a speaker in order to react appropriately. Theories of language acquisition posit that understanding others' communicative intentions is developmentally prior and causally related to the acquisition of meaningful language use (Tomasello, 2003), and experimental research confirms that infants around their first birthdays understand others' communicative acts in terms of underlying referential and social intentions (Tomasello et al., 2007). But how does this process work, and what are the sources of information that infants can use to infer meaning? Unravelling the kinds of cues infants pick up on when they make inferences about others' communicative acts is crucial to understanding the cognitive nature and build-up of the human capacity to infer others' communicative intentions.

One source of information infants heavily rely on when inferring others' communicative intentions is the preceding shared action context. For example, 18-month-olds imitate either the style or the outcome of an act depending on what of the two components had been introduced in the preceding action context (Southgate, Chevallier, & Csibra, 2009). Twelve- and 14 month-olds infer

referents of ambiguous requests (Liszkowski, Carpenter, & Tomasello, 2008; Moll, Richter, Carpenter, & Tomasello, 2008; Moll & Tomasello, 2007) and they infer reference to occluded (Behne et al., 2012) or misplaced referents (at 17 months; Southgate et al., 2010), depending on the information presented in the act-preceding action contexts. Infants also distinguish different social intentions underlying communicative acts. For example, they distinguish whether a point is meant to share interest in a referent or request it (Camaioni et al., 2004), or inform about its existence in a hiding game (Aureli et al., 2009; Behne et al., 2012), again based on the information from act-preceding action contexts and joint visual scenes (Liebal, Behne, Carpenter, & Tomasello, 2009; Liebal & Tomasello, 2009).

However, it is difficult to tell from these studies what exactly infants understand of the communicative act itself. On the one extreme the shared action context might be so routinized as to making a communicative act to elicit appropriate reactions superfluous. For example, as a regular of the university cafeteria the waiter anticipates my order and makes me a coffee and I anticipate his bill and place my coins on the cash tray, without need to instigate each step in the sequence through a communicative act. On the other extreme, a decisive, disambiguating preceding action context may be missing, for example when a new activity is initiated. I might go sailing with a novice colleague and exclaim "the sheet" and "that one, that rope" while pointing to it to resolve the reference, but without ever having sailed before, he may be puzzled as to what to do with it, or whether to be worried or join in

joy. Infants are novices on many tasks, and they face communication within underspecified shared action contexts quite often, especially since everyday communication is not as neatly preceded by clear-cut laboratory-style induced action contexts. Rather, infants are confronted with communicative acts that are embedded in more or less ambiguous or novel preceding shared action contexts. As adults, we can cope with some (though by far not all) of these instances in which the preceding action context alone would underspecify pragmatic meaning, because communicative acts are typically realized with several act-accompanying characteristics such as prosodic cues, which provide yet another source to pragmatic meaning (Pierrehumbert & Hirschberg, 1990). So, in the sailing example, if my attention-directing gesture to the sheet is accompanied by a questioning intonation or perhaps an open hand, it may be clearer that I want the sheet be handed over to me.

Thus, in cases of directing infants' attention through acts like pointing, as for example, when auntie comes home from work and points to a block for her nice who's building a tower, it could mean several things. But if aunty accompanies the act with an excited "Wow, nice", the infant might interpret it as auntie liking the block and thus look at it, or perhaps, pick it up to show it to her and share her interest in it. If, on the other hand, auntie distinctly utters with a falling tone of voice "that one", the infant might rather interpret it as auntie informing about that particular block, perhaps to continue building the tower with it. And perhaps, if auntie uses a questioning intonation and tilts the hand palm slightly up, infants might interpret

it as a request for the block and hand it over to her. Studies on language processing claim that speakers rely on various information types such as word meaning, syntax, social action context and world knowledge, to comprehend the message (Clark, 1996; Marslen-Wilson & Tyler, 1980; Tanenhaus & Trueswell, 1995). Naturalistic observation studies, however, cannot easily distinguish between situational information from the shared action contexts and accompanying characteristics of the acts, and 'in the wild' it is notoriously difficult to discern whether infants base their communicative inferences on the meaning of the lexical cues, on the prosodic and gestural accompanying characteristics of an act, on the preceding shared activities, or any combination of these factors (e.g., Esteve-Gibert, Liszkowski, & Prieto, in press). Experimental studies comparing infants' pragmatic understanding of different types of social intentions are scarce, and all manipulated the preceding action context (Aureli et al., 2009; Camaioni et al., 2004). In Camaioni et al. (2004), for instance, in a requestive pointing condition, an experimenter was playing with a toy that could be pulled apart in two pieces, then gave one of the two pieces to the child, and then expressed discomfort about not having that piece to keep on playing; when the experimenter then pointed toward the object in the infant's hands, the infant understood the requestive intention of the pointing gesture and gave back the piece to the experimenter. Instead, in a declarative pointing condition, a flashing light or a picture appeared at a distance behind the infant, and the experimenter pointed towards it; in this case, the child smiled or vocalized toward the stimulus, or reenacted what the stimulus did

while looking at the experimenter in an apparent attempt to share her interest. These studies thus did not address infants' use of act-accompanying features as one source of information to pragmatic meaning. Further, they confounded the distance and shared perceptual availability of the referent across the two conditions. Thus, it is currently unknown whether infants can infer pragmatic meanings of attention-directing acts based on accompanying characteristics alone, that is in the absence of disambiguating information from a shared action context and perceptual scene (for older children, see Liebal, Carpenter, & Tomasello, 2011).

Infants understand others' intentions and engage in shared activities towards the end of the first year of life, before they begin to direct others' attention in meaningful ways. Infants may thus have ample opportunity to learn about act-accompanying characteristics within these meaningful shared interactions, provided that parents consistently accompany their acts with distinct cues (Fernald, 1989; Koterba & Iverson, 2009). Like infants learn the referential meaning of words within joint engagement, it could well be that they also pick up on the pragmatic meaning of accompanying characteristics with which distinct acts are realized. For example, a recent semi-naturalistic study of 12-month-old infants in their home environments (Esteve-Gibert et al., in press) found that parents use distinct prosodic patterns and gesture hand shapes when drawing infants' attention to a referent in the course of different play formats. Parents mostly used high pitch range and index-finger pointing when sharing interest in exciting events; moderate pitch range and open-hand gesture when requesting objects; and narrow

pitch range and index-finger pointing when informing about hidden toys. However, in that study the play formats differed vastly in the type of preceding shared activities, and infants' response opportunities were not well controlled, so that it remained unclear whether infants understood parents' communicative acts appropriately and, in particular, whether they considered the accompanying characteristics of the acts as a source of information to the intended meaning.

To test infants' understanding of act-accompanying cues, we designed a new lab-based procedure in which we equated the preceding play activity, spatial layout, and response opportunities across conditions, and instructed parents to express one of three types of social intentions to their infants: expressive; requestive; informative (see Tomasello et al., 2007). If parents distinctly express their social intentions, we reasoned, they should accompany their attention-directing acts with rich information and thereby provide infants with additional cues to the intended meanings. In turn, if infants have acquired some abstracted understanding of accompanying characteristics as cues to pragmatic intentions, infants should react appropriately to these distinctly expressed meanings, even though the shared activity and perceptual co-presence would allow for various interpretations. That is, when parents express their interest in an object, infants should mostly share attention with them; when parents request an object, infants should rather offer it to them; and when parents provide information about a hidden item, infants should explore it. In Experiment 1 we pursued two main objectives. We tested whether infants would react

appropriately to communicative acts with different intended meanings even when the shared activity and perceptual scene within which these acts occurred was the same across conditions. We then analyzed parents' act-accompanying characteristics in terms of prosody and gesture shape to determine relevant differences in the information infants are exposed to other than the preceding shared action context. In Experiment 2 we used the same paradigm, albeit with trained experimenters, and controlled for the lexical content of speech, to exclude the possibility that 12-month-olds could perhaps base their responses on syntax or a semantic understanding of the lexical content alone.

5.2. Experiment 1

5.2.1. Method

5.2.1.1. Participants

Eighteen caregiver infant dyads participated in the study (9 girls). The mean age of the infants was 12 months, 17 days (range: 12 months, 7 days - 12 months, 28 days). Four additional caregiver-infants dyads were tested but excluded from the sample because of parental procedural error (N = 1) or because infants refused to participate (N = 3). All dyads were recruited from a Dutch database

of parents from a middle-size city in The Netherlands who expressed interest in participating in research with their child.

5.2.1.2. Set-up and materials

Two tables were arranged in an L-shape (see Figure 20). A child chair was attached to the table at the inner top side of the L. The caregiver's chair was placed 90º to the side of the child. Under the table and in front of the caregiver's chair, there was a small bench with two wooden toys on it. A black cardboard occluder was placed on the lower part of the L-shape table, behind the infant's side line, to hide the experimenter (E, henceforth) and stimuli. The cardboard had a small opening on the bottom in the middle. Two room dividers flanked the lower sides of the L and blocked the infant's view, so that infants could not see E during the experiment. One camera recorded the caregiver and infant. The camera was connected to a video screen that provided E behind the occluder with a full view of the scene.

A total of eight stimuli (cupcake cups) were used, one for each trial. The cups were all differently colored and looked appealing to the child. They were presented in front of the child on a black stick. The black stick had a round plate-like platform on its end onto which the cups were placed (see Figure 21). A round colored sticker was attached to the platform of the black stick, covered by the cup.

The same set-up and materials were used across the three conditions.

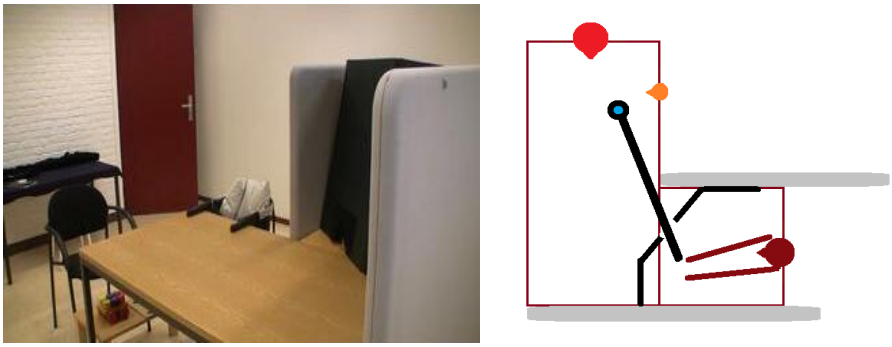

Figure 20. Set-up of the test room with two tables in an L-shape, the child chair on a corner, the caregiver's chair positioned at 90º, the black occluder hidden between two room dividers, and the experimenter hidden behind the occluder. Left, screen shot; right, schematic sketch of the set-up.

5.2.1.3. Procedure

Infants were randomly assigned to one of three conditions in which the caregivers were instructed to act in one of three ways, resulting in 6 caregiver-infant dyads per condition. There were 8 test trials per condition. The general procedure in the warm-up and test phase was identical in the three conditions.

Warm-up phase

In a separate room the caregiver, the infant, and E played with some toys for 5 minutes as a warm-up. Meanwhile, E explained the experiment to the caregiver in a general way. Then she gave specific instructions to the caregiver on what to do in the test phase. These instructions differed across conditions (see below). After the warm-up, all three went into the test room. E helped the caregiver to accommodate the infant in the infant seat, and instructed the caregiver to sit down in the caregiver's chair. E briefly reiterated the instructions again to the caregiver and then hid behind the occluder.

Play phase

In each trial, caregiver and infant played with the wooden toys about a minute, with the explicit instruction of having only one toy at a time on the table. The free-play was crucial to distract the infant before the adult directed the infant's attention towards the target object on the other end of the table. Otherwise, the infant could have focused on or manipulated the target object by chance irrespective of the adult's communication.

Test phase

The E made sure that the infant played with the caregiver, and did not attend to the occluder, and then protruded the upside-down

cupcake on the black stick inconspicuously through the opening of the occluder until it was 10 cm far from the child's seat (see Figure 21). There was a white mark behind the occluder and on the stick to signal how far E had to push the stick. E could monitor the distance on the camera and could adjust it depending on how far the child could reach: after the first trial, if E saw that 10 cm was too easy for the child to reach, the stick was left a bit further away, and vice-versa. Importantly, the distance between the caregiver and the cup was big enough to induce caregivers to use a deictic gesture to direct the infants' attention towards the cup. We chose not to have the target object in infants' view from the beginning of the trial to prevent infants from exploring or taking the cupcake cup before the caregiver actually directed the attention toward it.

Once the cup was placed in front of the infant and did not move anymore, the caregiver had to direct the infant's attention to the cup with an expressive, imperative, or informative motive (see Figure 21).

Expressive condition. Caregivers had been instructed to direct the infant's attention to the cup in order to share his/her interest about the cup with the child. The explicit instruction given to the caregiver was (in Dutch): *Deel je interesse voor de beker met je zoon/dochter. Gebruik gerust woorden of gebaren als u wilt. Het enige is dat u het object zelf niet aan mag raken* ('Share your interest about the cup with your son/daughter. Feel free to use words or gestures if you want to. The only thing is that you should not touch the object yourself').

Imperative condition. Caregivers were instructed to direct the infant's attention to the cup in order to get the child to give him/her the cup. The explicit instruction given to the caregiver was: *Vraag je zoon/dochter om jou de beker te geven. Gebruik gerust woorden of gebaren als u wilt. Het enige is dat u het object zelf niet aan mag raken* ('Ask your son/daughter to give you the cup. Feel free to use words or gestures if you want to. The only thing is that you should not touch the object yourself').

Informative condition. Caregivers were instructed to direct the infant's attention to the cup in order to inform the child that there was a sticker under it. The explicit instruction given to the caregiver was: *Informer je zoon/dochter dat er iets onder de beker verstopt is. Gebruik gerust woorden of gebaren als u wilt. Het enige is dat u het object zelf niet aan mag raken* ('Inform your son/daughter that there is something hidden under the cup. Feel free to use words or gestures if you want to. The only thing is that you should not touch the object yourself'').

Importantly, no explicit instruction was given to the caregiver on the gesture and speech strategies they had to use. E told the caregiver to direct the infants' attention towards the cup only once or twice per trial to prevent them to be very insistent. In total, the cup was in the infant's vision for 20 seconds. If the infant took the cup before the 20 seconds were over, the caregiver placed the cup back on the stick and E retracted the stick again. However, if the infant had not shown any reaction, E retracted the stick with the cup again behind the occluder. If the 20 seconds were over and the

infant was still playing with the cup, E shook the stick to signal the caregiver that the cup had to be placed on the stick again, and retrieved the cup. Then the play phase of the next trial started. Each trial involved a differently colored cup and sticker.

Figure 21. Screen shot during Experiment 1.

5.2.1.4. Data coding

The data was first coded in terms of the infant's behavior and then with regard to the caregiver's use of speech and gesture.

Infant behavior

Infant's behavior after each trial was coded using ELAN software (Lausberg & Sloetjes, 2009). Four categories were used: (a) *offering*

cup, when the child took the cup from the stick and gave it to the caregiver; (b) *attending cup*, when the child looked at the cup ostensively, pointed at it, or took it and played with it; (c) *attending sticker*, when the child took the cup off the stick and looked, pointed or played ostensively with the sticker or the black stick under the cup; (d) *no reaction*, when the child did not show any of these reactions within the 20 seconds during which the cup was placed in front of him/her. When more than one of these reactions was observed within the 20 seconds of a trial, the coder chose the primary reaction of the child, i.e. the reaction that was most salient, longest or not a consequence of incidentally discovering the cup or the sticker. Thus, if the child explored the cup for some seconds in order to then give it to the caregiver, the most salient behavior was *offering cup* and coded as such. Also, if the child took the cup, manipulated it for a couple of seconds but later left the cup apart to pay attention to the sticker and this second behavior lasted longer than the manipulation of the cup, it was coded as *attending sticker.* Finally, if the child attended the sticker because (s)he discovered it by chance after having manipulated the cup, this was considered an incidental behavior and the primary behavior was still considered to be *attending cup.*

Inter-rater reliability was conducted with 20% of the data of each condition (N = 26) by two independent coders who were unaware of the purpose of the study. The reliability in coding of the infants' behavior into offering cup, attending cup, or attending sticker, was 92.3%, Cohen's Kappa = 0.90.

Caregiver behavior

We coded the caregiver's use of gesture and speech in those trials in which infants complied with the instruction given by the caregivers, i.e. attending cup in the expressive condition, offering the cup in the imperative condition, attending the sticker in the informative condition in order to reveal the prosodic and gesture strategies that are successful in triggering the expected behaviors in children.

First, the gesture shape of the caregivers' deictic attention-directing gestures (either accompanied by speech or not) were annotated using ELAN software. The transcriber classified the caregiver's pointing gesture as *index-finger pointing gesture*, when the arm was extended and the index finger was directed to a specific location, or *whole-hand pointing gesture*, when the arm was extended and the hand was open and palm-up.

Second, the prosodic features of caregivers' speech (with or without accompanying gestures) were transcribed with the help of Dutch experts using Praat software (Boersma & Weenink, 2012). Within a trial, the utterance selected for the prosodic analyses was the one conveying the target intended meaning in each condition. When caregivers produced more than one utterance to convey the target intended meaning within a trial, they were all analyzed as two different data points for that trial. Three features were coded: the intonation pattern, the mean syllable duration, and the pitch range (see Figure 22). These three features were chosen on the basis of

previous research showing their relevance in early prosodic comprehension and production (e.g., D'Odorico & Franco, 1991; Kent & Murray, 1982; Papaeliou et al., 2002). Intonation pattern was annotated using the established ToDI system for the transcription of Dutch intonation (Gussenhoven, 2005). The first author, a trained phonologist (non-native Dutch) carried out a complete training in the ToDI transcription system following a courseware (Gussenhoven, Rietveld, Kerkhoff, & Terken, 2003) and participated in further individual sessions with Dutch-speaking ToDI experts for detailed guidance. The mean syllable duration was calculated by dividing the total duration of the sentence (in ms) into the number of syllables it contained. Pitch range of the utterance was coded by locating the maximum and minimum pitch points in the F0 line and subtracting the minimum to the maximum pitch point.

Inter-rater reliability was conducted with 10% of the data of each condition (N = 21) by two independent trained coders who were unaware of the purpose of the study. The reliability of ToDI intonation coding was 75.4%, Cohen's Kappa = 0.71[7], and 100% when coding gesture shape, Cohen's Kappa = 1.0.

[7] These results are consistent with the inter-transcriber agreement results found in studies on intonation transcription using the ToBI system (see Escudero, Aguilar, Vanrell & Prieto, 2012 for a review)

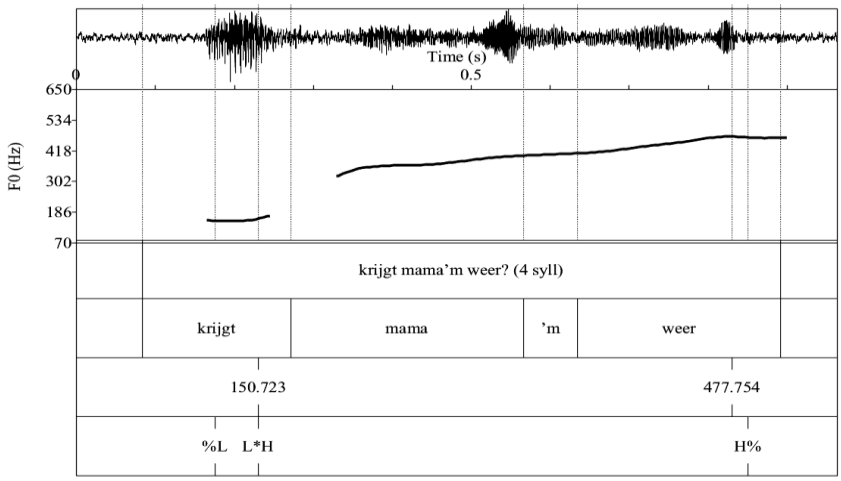

Figure 22. Example of the speech annotation in Praat of the sentence *krijgt mama'm weer?* ('does mummy get it again?'). The first tier shows the orthographic transcription and number of syllables of the utterance, the second tier shows the utterance word by word, third tier shows the values of minimum and maximum pitch in the F0 line in order to calculate the pitch range, and the fourth tier shows the transcription of the intonation in ToDI (Gussenhoven, 2005; Gussenhoven et al., 2003).

5.2.2. Results

a) Infants' behaviors

From the total amount of data (144 trials), 14 trials of 6 individual infants were excluded from the analysis because of parental procedural error (6 trials of 1 individual infant) and infants' refusal to participate (e.g., tearing the object or throwing it away; 8 trials of

5 individual infants). Thus, a total of 130 valid trials were used for subsequent analyses. The expressive condition yielded 91.7% valid trials, the imperative condition 97.9% valid trials, and the informative condition 81.3% valid trials.

A 3 (behavior: attending cup, offering cup, attending sticker) x 3 (condition: expressive, imperative, informative) repeated measures ANOVA revealed significant differences across behaviors ($F(1.661, 24.919) = 15.959$, $p < .001$) and an interaction between behavior and condition ($F(3.323, 24.919) = 8.235$, $p < .001$). To test our hypotheses we compared each behavior across conditions (see Figure 20). Infants attended the cup more often in the expressive than in the imperative or informative conditions ($F(1,10) = 7.960$, $p < .05$ and $F(1,10) = 13.159$, $p < .01$, respectively); they offered the cup more often in the imperative than in the expressive or informative conditions ($F(1,10) = 13.729$, $p < .01$ and $F(1,10) = 10.865$, $p < .01$, respectively); and they attended the sticker more often in the informative condition than in the expressive or imperative ones ($F(1,10) = 6.598$, $p < .05$ and $F(1,10) = 8.101$, $p < .05$, respectively).

Further analysis of the behaviors in each condition confirmed that the infants' behaviors were different in all conditions (expressive condition: $F(2,14) = 17.749$, $p < .001$; imperative condition: $F(2,14) = 4.673$, $p < .05$; informative condition: $F(2,14) = 6.029$, $p < .05$). In the expressive condition, infants attended the cup significantly more than they offered it or searched for the sticker (respectively $t(5)=7.826$, $p < .01$; $t(5)=4.909$, $p < .01$), with no differences

between the latter two ($t(5)=-.598$, $p = .576$). In the imperative condition, infants offered the cup and attended the cup more than they attended the sticker (respectively $t(5)=3.322$, $p < .05$; $t(5)=3.230$, $p < .05$), and they attended and offered the cup about equally often ($t(5)=.307$, $p = .771$). In the informative condition infants searched for the sticker marginally more than they offered the cup ($t(5)=-2.429$, $p = .059$), they attended the cup more than offered it ($t(5)=3.500$, $p < .05$), and they attended the sticker and the cup equally often ($t(5)=-.315$, $p = .765$).

Thus, infants attended the cup a lot in all conditions but across conditions they did this most in the expressive condition; they offered the cup across conditions most in the imperative one, and they attended the sticker across conditions most in the informative condition (see Figure 23).

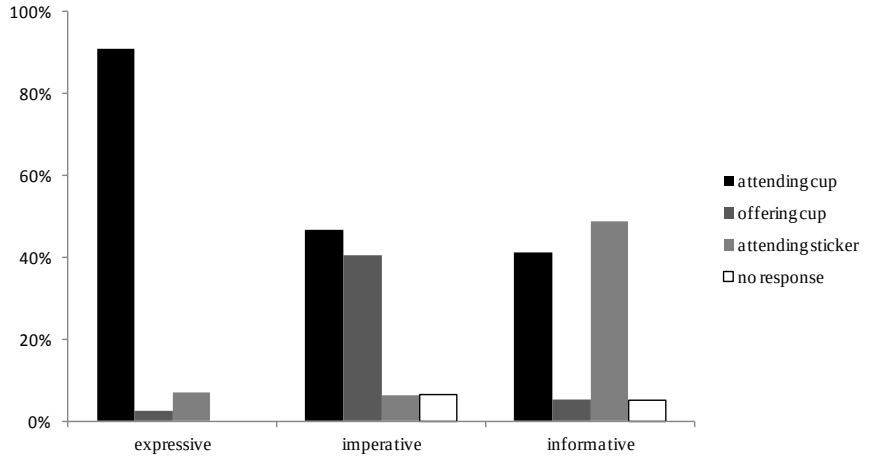

Figure 23. Percentage of children's behaviors across conditions in Experiment 1.

There was no evidence that infants learned over trials. There was no significant difference between the mean proportion of trials with the appropriate response in the first half of the trials (55.6%) compared to the second half of the trials (52.8%), paired-samples T test: $t(17) = 0.287$, $p = .777$. A 3 (condition: expressive, imperative, informative) x 2 (appropriate response: in the first half of the trials, in the second half of the trials) repeated measures ANOVA revealed no effect of condition on the amount of appropriate behaviors over trials ($F(1,15)=.204$, $p = .658$) and no interaction between condition and appropriate behavior ($F(2,15) = .663$, $p = .530$). We also analyzed how many of the children showed the expected behavior for each condition at least once during the experiment. First, all infants participating in the expressive condition attended the cup at least once during the experiment trial ($N = 6$). Second, in the imperative condition all infants offered the cup to the adults at least once during the experiment ($N = 6$). Third, in the informative condition all infants except for one attended to the sticker at least once during the experiment ($N = 5$) (see Table 12).

	Attending cup	*Offering cup*	*Attending*
Expressive	**6**	1	0
Imperative	6	**6**	3
Informative	5	1	**5**

Table 12. Number of children showing the three behaviors at least once in each condition.

b) Caregivers' patterns

The second aim of this experiment was to identify the specific gesture and speech strategies that adults used successfully to convey imperative, expressive, and informative intentions. Specifically, the purpose was to determine the gesture shape and prosodic cues that distinguish one intention from the other. From the total of valid trials (N = 130), in 67.7% (N = 88) the caregiver produced at least one combination of pointing and speech during the trial to direct the infant's attention, in 26.9 % (N = 35) caregivers used only speech strategies, and in the other 5.4% (N = 7) caregivers used a pointing gesture without speech. There was no relation between the pragmatic condition and the caregivers' use of a specific type of act ($\chi^2(4, N = 130) = 2.78$, $p = .596$) (see Table 13).

	Expressive	*Imperative*	*Informative*
G	1	4	2
G+S	32	32	24
S	13	10	12

Table 13. Number of each specific caregiver's act per condition.

The analysis of the caregivers' use of different gesture shapes across conditions revealed that caregivers always used index-finger pointing gestures during the expressive and informative conditions (100% of the cases in both conditions), while they always used

whole-hand pointing gesture during the imperative condition (100% of the cases), mostly with the hand palm oriented up.

The analysis of the caregivers' use of prosodic cues in the pointing-speech combinations across conditions took into account three features: (1) the intonation contour, using ToDI (Gussenhoven, 2005; Gussenhoven et al., 2003, but not taking into account initial boundary tones to reduce variance), (2) the mean syllable duration, in milliseconds, and (3) pitch range, or distance in semitones between the highest and the lowest F0 locations in the pitch contour.

The first sub-analysis on intonation showed that four contours were the most commonly observed (representing 75% of the total utterances, see Figure 24). From these four main contours, the most frequent one was the fall (H*L L%, 40.4%, N = 84), followed by the fall-rise (H*L H%, 14.4%, N = 30), the half-completed fall (H*L, 12.1%, N = 25), and the low rise (L*H H%, 8.2, N = 17). A repeated-measures ANOVA with the four most frequent intonation contours (fall, fall-rise, half-completed fall, low rise) as dependent variable and condition as between-subjects variable (3 levels: expressive, imperative, informative) revealed a main effect of intonation (F(3,45)=31,609, p < .001) and an interaction between intonation and condition (F(6,45)=3.245, p <.01) (see Figure 24).

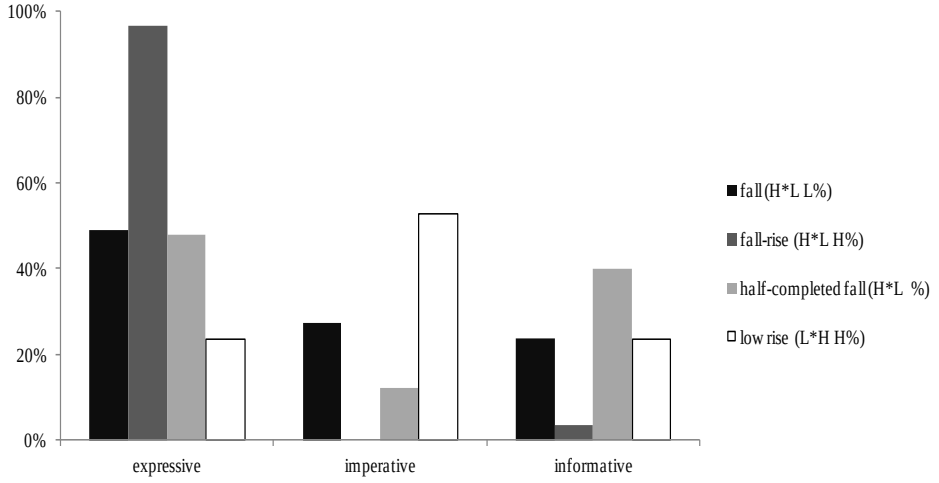

Figure 24. Proportions of the four most frequent intonation contours displayed across conditions, transcribed in ToDI (Gussenhoven, 2005; Gussenhoven et al., 2003).

Subsequent pair-wise comparisons showed that the fall contour (H*L L%, see Figure 25 top panel) occurred significantly more often in the expressive than in the informative condition ($F(1,12) = 10.804$, $p < .01$), all other comparisons with this contour being non-significant. The fall-rise contour (H*L H%, see Figure 25 bottom panel) occurred significantly more often in the expressive condition than in the imperative or informative conditions (respectively, $F(1,12)=5.840$, $p < .05$ and $F(1,12)=4.793$, $p < .05$). The half-complete fall and the low rise contours (H*L and L*H H%, respectively) did not differ significantly across conditions (respectively, $F(2,17) = .745$, $p = .492$ and $F(2,17) = .027$, $p = .473$).

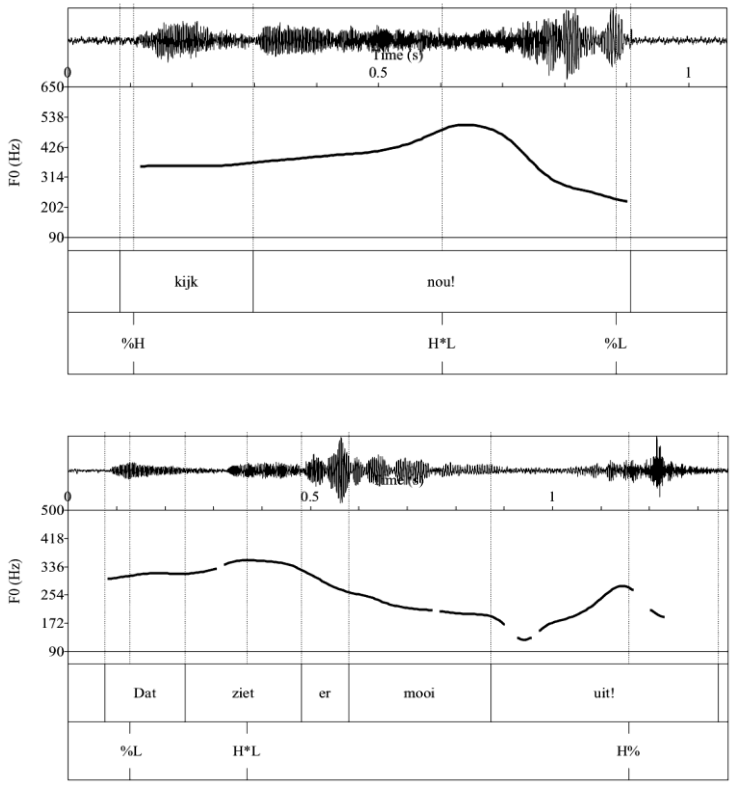

Figure 25. Top panel, example of the most frequent intonation. the fall (H*L L%); bottom panel, example of the fall-rise contour (H*L H%) that was observed significantly more often in the expressive condition.

As for the second sub-analysis on the mean syllable duration, a one-way ANOVA analysis revealed that the fixed factor "condition" had a significant effect on the dependent variable "mean syllable duration" ($F(2, 207)=34.063$, $p < .001$), and follow-up comparisons showed that the mean syllable duration differed significantly between the expressive and the imperative condition ($F(1,180) = 37.824$, $p < .001$, $\eta^2 = .174$) and between the expressive and the

informative condition ($F(1,170) = 19.430$, $p < .001$, $\eta^2 = .103$), but not between the imperative and the informative condition ($F(1,78) = 3.078$, $p < .083$, $\eta^2 = .038$). The left panel of Figure 26 shows that the mean syllable duration (in ms) was longer in the expressive condition compared to the other conditions, meaning that caregivers spoke more slowly when they shared their interest for the object with the child than when requesting an object or informing about its presence.

As for the third sub-analysis on the pitch range, a one-way ANOVA analysis showed that the fixed factor "condition" also affected significantly the dependent variable "pitch range" ($F(2, 208)=4.957$, $p <. 01$). As shown in Figure 26 (right panel), caregivers used a wider pitch range in the expressive and imperative conditions than in the informative condition. Follow-up comparisons revealed a statistically significant difference between pitch range in the expressive condition compared to the informative condition ($F(1,170) = 6.710$, $p < .01$, $\eta^2 = .038$). By contrast, the pitch range in the expressive condition was similar to the imperative condition ($F(1,180) = 2.651$, $p = .105$, $\eta^2 = .015$), and the imperative and the informative conditions did not differ significantly from one another ($F(1,78) = 1.154$, $p < .286$, $\eta^2 = .015$).

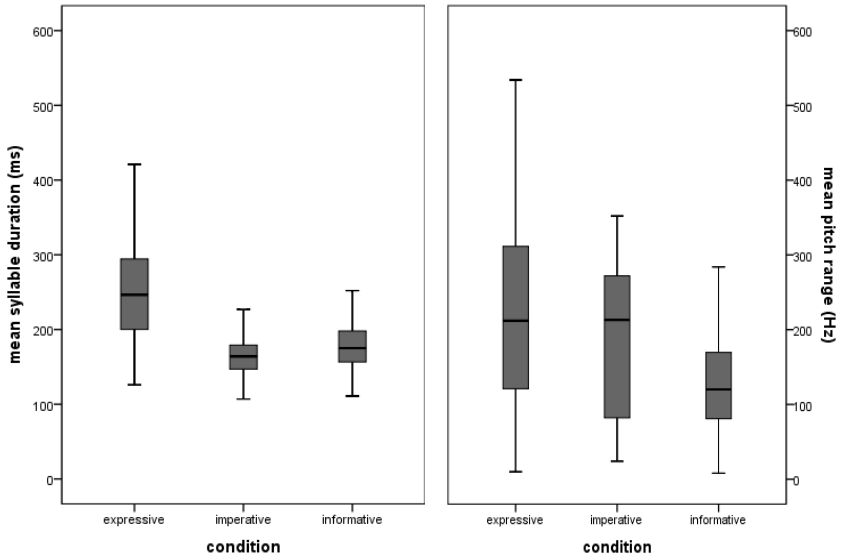

Figure 26. Left panel, box plots of the mean syllable duration (in milliseconds) in the three conditions (expressive, imperative, and informative). Right panel, box plots of the mean pitch range (in Hertz) displayed by the caregivers as a function of the condition.

5.2.3. Discussion

In Experiment 1 caregivers were instructed to direct infants' attention towards a cup with the underlying intention of asking for the cup, sharing their interest about the cup, or informing about a hidden sticker under the cup. Twelve-month-old infants interpreted the social intentions underlying these attention-directing acts mostly appropriately and accordingly offered the cup, shared interest in the cup, or searched for something under the cup. What sources of information did infants use to comprehend the underlying social intentions of parents' attention-directing acts?

The experiment controlled across conditions for information stemming from preceding action contexts and perceptual co-presence, two forms of 'common ground' which have been shown to influence infants' behaviors and presumably their interpretations of communicative acts. Across all experimental conditions infants played with their parent with the same toys; their attention was not focused on the target object before the communicative act; and the spatial layout, seating arrangement and response period was the exact same across conditions. Therefore we can exclude that forms of common ground like preceding action contexts and perceptual co-presence differentially influenced infants' behavior across the three conditions. One could argue that infants differentially learnt over trials and accumulated some form of common ground over repeated trials. However, parents were instructed to always put the cup back on the stick, which would be a rather unexpected reaction to a fulfilled communicative act in the requestive and informative conditions. More importantly, a direct test of learning over trials revealed no increase of target behaviors, so we can exclude that information from repeated trials differentially influenced infants' responses in each condition.

The experimental results rather show that infants interpreted the underlying social intention of the attention-directing acts based on concurrent information emanating from the act itself. Our analyses of caregivers' communicative acts provide evidence for natural differences in the way caregivers express their reasons for directing infants' attention to an object. In line with recent findings from home observations in semi-structured, albeit less controlled

situations (Esteve-Gibert et al., in press), caregivers used unique patterns of gesture shape and phonetic cues of prosody across conditions: in the expressive condition syllables were long, utterances had a wide pitch range, and the pointing gesture had an index-finger pointing shape; in the imperative condition syllables were short, utterances had a reasonably wide pitch range, and the pointing gesture had a whole-hand shape; in the informative condition, syllables were short, utterances had a narrow pitch range, and the pointing gesture had an index-finger pointing shape. Caregivers used most often the fall intonation (H*L L%) across conditions, although statistical analyses revealed that it occurred more often in the expressive than in the informative condition. Also, the fall-rise contour (H*L H%) was particularly characteristic of the expressive condition.

Thus, parents expressed the distinct social intentions underlying their attention-directing acts each with a unique pattern of gesture shape and prosody. The fact that expressive and informative pointing gesture shared the same hand shape feature is consistent with the underlying nature of these social motives: while expressive and informative pointing gestures are both "declarative" in the sense that they aim primarily at directing attention toward something to signal an epistemic state, imperative pointing gestures primarily signal a motivational state (i.e. desire to obtain an object).

The meaning of prosody is usually established as a function of context within which it is used. In our experiment, however, adults were instructed to convey different intentions, and so the prosodic

realization was the dependent (not independent) variable. The most common intonation contour started high and then fell, presumably reflecting the attention-drawing element common to each condition. The expressive condition had the most varied contours; widest pitch range and longest syllable durations, reminiscent of infant-directed speech characteristics which have also been related to affective talk generally (Bryant & Barrett, 2007; Trainor, Austin, & Desjardins, 2000), a pattern which corresponds to our expressive condition of sharing an affective state with the infant. In the informative condition, in contrast, contours varied least and syllables were short, presumably reflecting the less emotion-laden content, highlighting the informational element. The imperative condition was also less affective compared to the expressive condition, however its unique contour had a rising shape, which is typically associated with questions for information or help in Dutch (Haan, van Heuven, Pacilly, & van Bezooijen, 1997). The upward-oriented palm gesture then provided a strong cue to the request of obtaining the object rather than conveying relevant new information.

Differences in prosody and hand shape thus distinguished caregivers' motives of pointing gestures. Based on our design and findings we can exclude that infants based their responses selectively on information stemming from different preceding action contexts or other perceptual aspects of the situation or learning over trials. Our interpretation of the results is that by 12 months infants have an understanding of act-accompanying characteristics, like prosody and gesture shape, which provides them with additional clues to parents' pragmatic intentions when

information from other sources of a common ground cannot sufficiently disambiguate the intended meaning. However, although 12-month-olds are prelexical in their own communication, and presumably have yet little context-free semantic understanding of labels (but see Parise & Csibra, 2012) or verb constructions, we cannot conclusively exclude that infants understood parents' intentions based on the lexical information of parents' speech. In order to control for the lexical cues, and in an attempt to replicate Experiment 1, we conducted Experiment 2, in which we substituted parents with experimenters to keep both the social contextual information and the lexical cues the same across pragmatic conditions.

5.3. Experiment 2

The aim of Experiment 2 was to replicate findings from Experiment 1 while controlling for the lexical information in the speech. We trained Experimenters on the caregivers' strategies identified in Experiment 1 regarding the hand shape and prosodic cues and had them express the same lexical content across conditions. Otherwise the paradigm was the same as in Experiment 1. Based on the results of Experiment 1 our predictions were that 1) in the expressive condition infants attended the cup more so than in the other two conditions; 2) in the imperative condition infants offered the cup more so than in the other two conditions; 3) in the informative

condition infants explored the cup/sticker more so than in the other two conditions.

5.3.1. Method

5.3.1.1. Participants

Thirty 12-month-old infants participated in the study (9 girls). None of them had participated in Experiment 1. The infants' mean age was 12 months, 12 days (range: 12 months, 3 days - 12 months, 26 days). Six additional infants were tested but excluded from the sample because they became fussy in more than half of the trials (*N* = 3), they did not want to play (*N* = 2), or because of mother interference (*N* = 2). All infants were recruited from a Dutch database of parents from a middle-size city in The Netherlands who expressed interest in participating in research with their child.

5.3.1.2. Set-up and materials

The set-up in Experiment 2 was identical to Experiment 1, with the only differences that (1) an experimenter sat on the place where the parent had previously sat, and that (2) there was a chair behind the infant on which parents sat during the experiment (see Figure 27).

5.3.1.3. Procedure

Infants were randomly assigned to one of three conditions resulting in 10 caregiver-infant dyads per condition. The procedure was the same as in Experiment 1 except that (1) an experimenter drew the infants' attention towards the objects; (2) specific gesture and prosodic strategies were used in each condition, and (3) the same lexical information was used across conditions. There were three experimenters each testing the same amount of infants in each of the three conditions.

Like parents, the experimenters had been instructed to convey three different types of social intentions. However, the experimenters were trained in two training sessions how to produce the same consistent gesture-prosodic strategies, which were modelled after caregivers' strategies that had triggered the expected infants' behaviors in Experiment 1 (see Figure 27). The training sessions were administered by the first author and consisted of watching a video with the target gesture-prosodic strategies and practicing how to produce them. Also, before testing the infants, the gesture-prosodic strategies were repeated to ensure that the experimenters remember them correctly. Finally, to reduce a potential effect of experimenter in the infants' behavior, all experimenters tested the same number of infants per condition. Further, the first author of the study was hidden behind the occluder during all the trials to check whether the experimenter performed the stimuli properly and, if necessary, provide feedback. We also controlled the body posture so that it did not differ across pragmatic conditions: the

experimenter was slightly moving forward during the pointing gesture and quickly recovered her initial state. The gaze patterns were also controlled: the experimenter alternated gaze twice between the cup and the infant during the attention-directing act and after that she fixed her gaze toward the infant for the rest of the trial. In all three conditions, E produced a pointing gesture with the same sentence and the same intonation contour. She said (in Dutch): "*Hey! Die! Die!"* ('Hey! This! This!') with a falling intonation contour (H*L L%). The specific gesture-prosodic strategies in each condition were the following:

Imperative condition. E produced a whole-hand pointing gesture, palm tilted slightly upwards. She used a wide pitch range and short syllables.

Expressive condition. E produced an index-finger pointing gesture. She used a wide pitch range and long syllables.

Informative condition. E produced an index-finger pointing gesture. She used a narrow pitch range and short syllables.

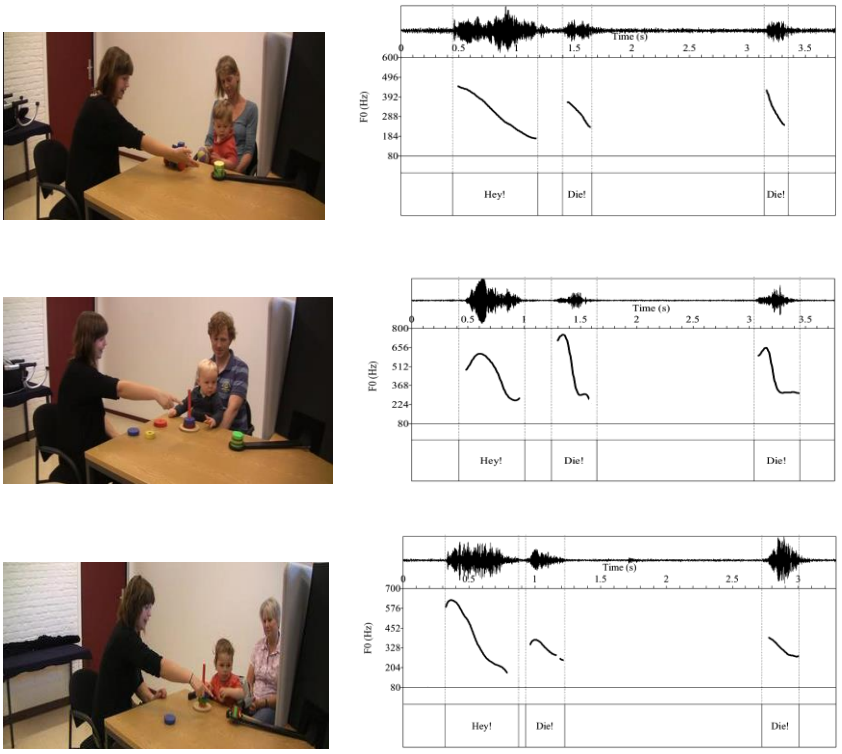

Figure 27. Stimuli used in Experiment 2. Top panel, stimuli in the imperative condition: whole-hand pointing gesture accompanied by a falling intonation contour with short syllables and wide pitch range. Middle panel, stimuli in the expressive condition: index-finger pointing gesture accompanied by a falling intonation contour with longer syllables and wide pitch range. Bottom panel, stimuli in the informative condition: index-finger pointing gesture accompanied by a falling intonation contour with shorter syllables and narrow pitch range.

The experimenter produced the pointing gesture accompanied by speech only once and then returned to her initial position, without interacting with the child during the rest of the trial except for the cases when the child offered the cup to them. In those cases, the experimenter took the cup affirmatively and then placed it back on

the stick. As for facial gestures, experimenters were told to have a smiling face during the expressive condition and a friendly but slightly less enthusiastic facial expression during the imperative and informative conditions in order not to act odd or violate the naturally occurring expression of the expressive motive.

5.3.1.4. Data coding

The data was coded as in Experiment 1. A manipulation check was conducted with 100% of our data to check for potential differences in Experimenters' behaviors between conditions in terms of body posture, facial gestures, and gaze alternations. In 100% of the trials the experimenter produced the trained body posture and gaze alternations. In 89% of the trials the experimenter produced the trained facial expression. Inter-rater reliability was conducted as in Experiment 1 with 20% of the data from each condition (N = 47) by two independent coders who were unaware of the purpose of the study. The reliability in coding of the infants' behavior was 92.4%, Cohen's Kappa = 0.90.

5.3.2. Results

From the total amount of data (240 trials), 2 trials were excluded from the analysis because of experimenter error, so a total of 238 valid trials were used for the analyses. Figure 28 shows an overview of the reactions produced by the child across conditions. A 3

(behavior: attending cup, offering cup, attending sticker) x 3 (condition: expressive, imperative, informative) repeated measures ANOVA revealed a main effect of behavior ($F(2,26) = 45.666$, $p < .001$, $\eta^2 = .778$) and an interaction between behavior and condition ($F(4,52) = 5.444$, $p < .01$, $\eta^2 = .295$). Follow-up one-way ANOVAs confirmed that each of the three behaviors occurred at significantly different rates within each condition (attending cup: $F(2,27)=7.195$, $p < .01$, $\eta^2 = .348$; offering cup: $F(2,27)=3.828$, $p < .05$, $\eta^2 = .221$; attending sticker: $F(2,27)=6.688$, $p < .01$, $\eta^2 = .331$). Following our predictions and results of Experiment 1 we conducted three planned contrasts to test for each behavior whether it followed the predicted pattern across the three conditions. The first planned contrast revealed that infants attended the cup significantly more often in the expressive condition than in the other two conditions ($t(27)=3.118$, $p < .01$). The second planned contrast revealed that infants offered the cup to the adult significantly more often in the imperative condition than in the other two conditions ($t(27) = 2.502$, $p < .05$). The third planned contrast revealed that infants attended to the sticker significantly more often in the informative condition than in the other two conditions ($t(27) = 3.657$, $p < .001$).

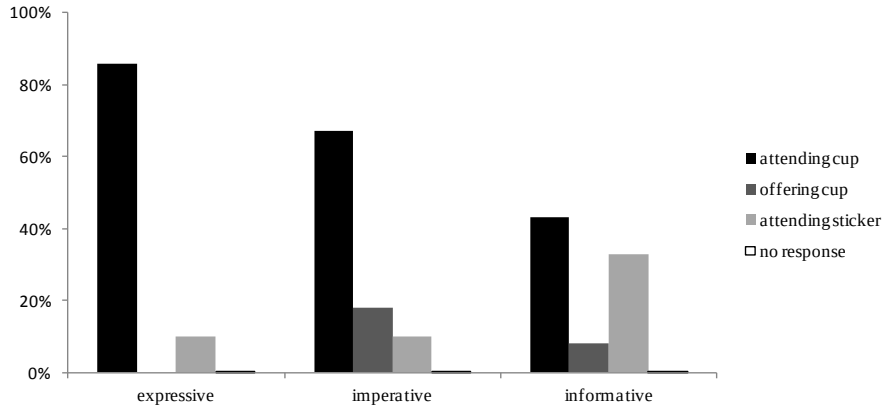

Figure 28. Percentage of children's reactions across conditions in Experiment 2.

There was no evidence that infants learned over trials. There was no significant difference between the mean proportion of trials with the appropriate response in the first half of the trials (45%) compared to the second half of the trials (47.5%), paired-samples T test: $t(29) = 0.462$, $p = .647$. A 3 (condition: expressive, imperative, informative) x 2 (appropriate response: in the first half of the trials, in the second half of the trials) repeated measures ANOVA revealed that the appropriate responses in the first half of the trials compared to the second half of the trials did not vary ($F(1,27) = .591$, $p = .449$) and no interaction between appropriate responses in the first and second trials and condition ($F(2,27) = .732$, $p = 490$). We also calculated the number of children showing the expected behavior at least once during the experiment to be sure that our results do not come from only one child being better than the others. First, all infants participating in the expressive condition attended the cup at least in one during the experiment trial ($N = 10$). Second, in the

imperative condition more than half of the infants offered the cup to the experimenter at least once during the experiment (N = 6). Third, in the informative condition all infants except for one attended the sticker at least once during the experiment (N = 9) (see Table 14).

	Attending cup	*Offering cup*	*Attending*
Expressive	**10**	0	5
Imperative	10	**6**	4
Informative	10	4	**9**

Table 14. Number of children showing the three behaviors at least once in each condition.

5.3.3. Discussion

Experiment 2 replicated the main findings from Experiment 1, while it controlled for the lexical information conveyed across conditions, thus ruling out the possible alternative interpretation of Experiment 1 that infants' reacted appropriately based on a lexical-semantic understanding. We tested infants in the three situations that were exactly the same in terms of social context and lexical information: the set-up and procedure was invariable across conditions and the experimenters always used the same lexical cues when directing the infants' attention towards the object. As in Experiment 1 our analyses of learning over trials revealed no change in behavior when comparing first half of the trials and second half of the trials across conditions, thus excluding the

possibility that infants' used differently accumulating information in the different conditions. The only available sources of information that we manipulated experimentally were the gesture shape and prosodic cues of pitch range and syllable duration (contour was constant, too). Thus, infants' different interpretations of the attention-directing pointing gestures were driven neither by differences in preceding action context, perceptual co-presence, nor the lexical content of speech. Instead, the experimental results reveal that infants reacted appropriately based on the shape of the pointing gesture combined with the prosodic cues of duration and pitch range.

Apart from the replicable condition differences, the current Experiment 2 also revealed that sharing attention was the most frequent behavior across conditions. One possibility is that sharing attention is a default interpretation for referential acts to new things, perhaps because the relevance of attention is guided bottom-up (gloss: 'oh, something new'). In the absence of pre-established common ground, it is presumably easiest and most natural for infants to interpret a pointing gesture as an expression of affective interest (Liebal & Tomasello, 2009), because this motive is rooted in (non-referential) communicative exchanges that develop soon after birth (Trevarthen, 1979). Instead, the interpretation of an informative and imperative pointing gesture (attending the sticker and offering the cup, respectively) depends on the understanding of the others' goals. The relevance is thus rather top-down guided, presumably requiring more cognitive effort to infer others' goals. In the absence of pre-established common ground this should be more

difficult; and it should require deeper processing of the style of the act. This could explain, among others, why the appropriate response rates in the imperative and informative conditions were overall a bit lower than in the expressive condition, suggesting that imperative and informative pointing gestures rely more on preceding common ground and social context than expressive pointing acts.

5.4. General discussion

Pragmatic accounts of development suggest that the acquisition of language and, perhaps, the emergence of social understanding more generally, are rooted in early social interactions and communicative exchanges. Accordingly, one endeavor has been to show that infants communicate meaningfully before they engage in earnest in verbal communication and explicit theory-of-mind reasoning (Liszkowski, 2013). From a cognitive point of view, the question is how infants do that. On what information do infants base their understanding of others' communicative acts, and specifically their understanding of others' social intentions? The few available studies on social intention comprehension all manipulated shared action contexts or perceptual co-presence in scenes that preceded the to-be-interpreted communicative act (Aureli et al., 2009; Behne et al., 2012; Camaioni et al., 2004; Liebal et al., 2009). The style of the communicative act was held constant and usually arbitrary in those studies to allow for different interpretations. Accordingly, the picture that has emerged from the previous findings is that infants

interpret pragmatic intentions underlying communicative acts based on the available information from preceding action contexts and perceptual co-presence.

We have argued, however, that this information may not always be readily apparent in everyday discourse with prelinguistic infants. Further, the lab-based studies provided extensive preceding action contexts, which may have yielded appropriate reactions without requiring a communicative act at all, thus questing how much infants understand of a communicative act itself. The current study provides experimentally controlled evidence that infants at 12 months of age have a pragmatic understanding of the three main types of social intentions underlying communicative acts (expressive, imperative, and informative intentions; Tomasello et al., 2007). Beyond previous findings, the current study shows that in the absence of disambiguating information from preceding action contexts and perceptual co-presence, infants rely on information that stems from the varying kinds of characteristics accompanying an act. This shows that by 12 months of age infants' cognitive system supporting mutual understanding does not process shared action contexts alone but also other kinds of input, like the form and style of acts.

The main kinds of cues we investigated were prosody and gesture shape. Previous findings (Esteve-Gibert et al., in press), and our internal validation of parents' behavior in Experiment 1, revealed that caregivers express types of social intentions differently on these behavioral dimensions. Our results show that infants are sensitive to

that when having to infer the others' intention. Traditional accounts on the acquisition of the pragmatics of prosody, for instance, had claimed that the comprehension of meanings conveyed through prosody is acquired relatively late and after children had actually used these features in production (Cutler & Swinney, 1987). Our results, together with other recent findings (Sakkalou & Gattis, 2012), point at an early development of the relation between prosody and pragmatics when this prosody is combined with other act-accompanying features such as gesture. Future studies could further refine the kinds and the scope of accompanying characteristics to which infants are sensitive. For example, it is conceivable that there are some natural, perhaps phylogenetically primary, culturally shared accompaniments in the emotional domain that signal e.g. approach or avoidance, as some social referencing studies suggest (Moses, Baldwin, Rosicky, & Tidball, 2001; Vaish & Striano, 2004). In adult communication, however, act-accompanying characteristics like prosody (Pierrehumbert & Hirschberg, 1990) and gesture shape (Kendon, 2004) are part and parcel of culturally shaped language use and acquired ontogenetically. The current study reveals that the pragmatic understanding of accompanying prosodic and gestural characteristics emerges prior to lexical-semantic cues. The cognitive skills for this kind of pre-lexical form-meaning mapping could well be related to the extraction of form invariances as required for word learning.

Infants likely first learn about act-accompanying characteristics in the first year of life through statistical co-occurrences with acts that

are embedded in extensively shared, meaningful action contexts, like rituals and routines. For example, at four months infants already anticipate others' actions in known routines (e.g., being picked up; (Reddy, Markova, & Wallot, 2013). These kind of interpersonal activities are accompanied by rich additional cues in prosody, gesture, and posture providing infants with ample opportunity to learn about act-accompanying characteristics and how they map into meaning. In this respect, infants' understanding of accompanying characteristics is rooted in simpler forms of action understanding. It should be seen as a developmental social-cognitive achievement of the first year of life, not just a precursor to meaningful language use.

6. GENERAL DISCUSSION AND CONCLUSIONS

6.1. Summary of findings

The goal of this dissertation was to investigate how infants integrate prosodic and gestural cues from a temporal point of view and how they use this integration for intentional communication. Four independent studies were presented, each one in a different chapter. The first two studies focused on the development of the temporal alignment between prosody and gesture in early infancy (Chapters 2 and 3). The last two studies focused on how young infants use the two modalities in an integrated way to communicate intentionally before producing their first words (Chapters 4 and 5).

Regarding the temporal alignment between prosody and gesture in early infancy, two main results were obtained. First, in Chapter 2 we found that infants produced pointing gestures in combination with speech from the beginning of the period analyzed (i.e., 11 months of age), but that most of the combinations started to occur when infants were 15 months of age. We also found that infants temporally aligned prosody and gesture in a fine-grained way from the beginning of pointing-speech combination production, since their gestures started before the corresponding vocalizations, gesture strokes started along with the beginning of the accented syllables, and gesture apexes occurred before the end of that accented syllable. Second, in Chapter 3 we found that the infants' sensitivity to the multimodal temporal alignment occurred well before they actually produced temporally aligned gesture-speech

combinations, i.e., when the infants were 9 months old, since at that age they were able to distinguish between properly aligned stimuli (in which the gesture prominence coincided with the prosodic prominence) and non-properly aligned stimuli (in which the gesture prominence did not coincide with the prosodic prominence).

With respect to the use of the integration between gesture and prosody for intentional communication, again two main results were obtained. First, in Chapter 4 we found that prosodic patterns (usually accompanied by gestures) were used by pre-lexical infants to signal intentionality in their vocalizations. Specifically, pitch range and duration values were found to distinguish between speech acts such as requests, statements, responses, expressions of satisfaction, and expressions of discontent. Second, in Chapter 5 we showed that infants were able to use prosodic cues (intonation, pitch range, or speech rate) and gesture shape (whole-hand or index-finger in pointing gestures) patterns to infer the intentionality of an action directed at them when the preceding shared action context did not give them enough relevant information. As a whole, these two chapters showed that infants not only interpreted or conveyed pragmatic intentions in communicative acts on the basis of the information in the preceding action contexts, but they also used linguistic cues like prosody and communicative features such as hand shape to do it so.

In the next two sections I will discuss these findings in relation to the previous literature and show how they contribute to the existing body of research, first, with regard to the temporal integration of

gesture and speech and second, with regard to the emergence of intentional communication.

6.2. The development of the temporal integration of prosody and gestures

Research on the integration of gesture and speech has proposed that gesture and speech form an integrated system in human communication (Kelly et al., 2010; Kendon, 1980; McNeill, 1992). One of the reasons used to support this claim is the fact that gesture and speech are aligned from a temporal point of view. In everyday conversations, gestures are mostly combined with speech, and when combined, their prominences tend to co-occur. Although the prominences have been understood slightly differently in studies of gesture and speech integration, most researchers would agree that stressed and accented syllables are the key anchoring points in speech (e.g., De Ruiter, 2000; Nobe, 1996; Rochet-Capellan et al 2008; Rusiewicz, 2010; Levelt et al., Loehr, 2012; Yasinnik et al., 2003). In gestures, the prominence is generally taken to be the gesture stroke (the interval of time involving the greatest physical effort and which carries the linguistic information in the gesture; Krahmer & Swerts, 2007; Rochet-Capellan et al., 2008), or the gesture apex (the point, not the interval, of greatest effort in the gesture; De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Loehr, 2012). Despite these slightly different criteria, very few researchers now doubt that gesture and prosodic prominences are temporally aligned in speech-accompanying gestures.

The first two studies described in this dissertation showed that the temporal alignment of gesture and prosodic prominences develops very early in language and cognitive development. Specifically, these studies showed that the development of the multimodal temporal alignment is grounded on the development of the rhythmic structure of speech. Nine-month-old infants have the ability to perceive the stress patterns of their target language (Höhle et al., 2009; Jusczyk et al., 1993; Nazzi et al., 1998; Pons & Bosch, 2010; Skoruppa et al., 2009, 2013), and our findings showed that already at that stage they are sensitive to the alignment between the prominence in pointing gestures and the prosodic prominence in speech. In production, infants at the one-word stage develop the stress patterning of their target language (Behrens & Gut, 2005; Davis et al., 2000; Snow, 2006; Vihman et al., 1998), and our results show that at that age they not only start combining most of their communicative gestures with speech but also temporally align the two modalities when they combine them. We hypothesize that the development of the rhythmic structure of speech is crucial for infants to anchor the prominence in gesture.

In sum, two main conclusions can be drawn. First, the development of prosodic-gesture alignment is related to the development of prosodic patterns. As soon as infants are sensitive to the timing of the prosodic prominence they are also aware that the gesture prominence of a speech-accompanying gesture must coincide with it. In addition, as soon as infants use the acoustic cues of prominence to produce the stress patterns of their language, they combine communicative gestures and speech with their respective

prominences properly aligned. Second, and crucially for research on the integration of gesture and speech, the gesture system and speech system are integrated very early on in language and cognitive development, at least as early as 9 months of age. As a result, these findings support the accounts suggesting that gesture and speech form a single system in human communication (De Ruiter, 2000; Kelly et al., 2010; Kita, 2003; McNeill, 2005).

Future research should provide a complete picture of the development of the temporal alignment between gesture and speech by investigating other types of communicative gestures and the neurocognitive basis of this alignment. As far as we know no research has investigated how children temporally align iconic gestures with the corresponding speech when they start producing them around 3 years of age. Similarly, gestures such as manual beats or head nods that appear during discourse develop later in the child's communicative development (around 4 or 5 years of age), and thus far no research has been done to see how infants align such gestures with co-occurring speech. In addition, it is essential to investigate the way temporal alignment interacts with semantic and pragmatic alignment in young children. Some studies report that adults can produce gestures related with speech in which the two modalities are not temporally aligned if the discourse situation or emotive implications requires this (Bergmann et al., 2011; Esteve-Gibert et al., 2014; González-Fuente et al., 2014). We believe that further research on other types of gestures and the neurocognitive basis of integration, together with the results presented in this dissertation, will reveal whether the alignment between gesture and

speech is something related to perceptual and motor abilities, or whether it also has to do with language comprehension and production.

6.3. The integration of prosody and gestures in early intentional communication

There is general consensus that prosody and gesture convey intentional meaning in adult speech. Speakers use prosody to identify sentence type (for instance, an interrogative or a declarative sentence), to structure information in speech (for instance, to distinguish between topic and focus), to express emotions, and to communicate meanings such as evidentiality and epistemicity. And speakers use gestures for deictic purposes with a related speech act (a pointing gesture with an imperative, expressive, or informative intention), for representational purposes (iconic or metaphoric gestures), or to help organize the information in discourse (beat gestures).

Previous literature on communicative development stated has established that the first sign of infants becoming intentional agents is their ability to produce and comprehend pointing gestures. On the one hand, infants start producing communicative pointing gestures between 8-12 months of age (Bates et al., 1979), and 12-month-old infants are able to produce pointing gestures with various social intentions, namely requesting an object from an adult (imperative intention), sharing interest in an object with the adult (expressive

intention), and signaling the importance of information to the adult (informative intention) (Liszkowski et al., 2004, 2006; Liszkowski, 2005). On the other hand, 12- to 14-month-old infants have been found to comprehend the intentional meaning behind pointing gestures (Aureli et al., 2009; Behne et al., 2012; Camaioni et al., 2004). These studies showed that infants infer the intention of communicative pointing gestures because they rely on the information available in the preceding social action context in which the pointing gesture occurs. For instance, if the infant is playing with an object that is important for the game and someone suddenly moves that object out of the infant's reach, the pointing gesture that the infant might produce will be understood as having an imperative meaning. Also, if the social context indicates that the adult needs an object and then subsequently the adult points towards that object, the infant understands the imperative meaning of the pointing gesture.

This dissertation has shown that prosody, and not only pointing gestures, is a sign of the infant becoming an intentional agent. The notion that prosody is part of the grammar is at this point beyond doubt. Prosody identifies sentence types, helps structure information in the discourse, and is used to express emotions and pragmatic meanings. Research on the development of the pragmatic uses of prosody had demonstrated that infants use prosody to distinguish intentional from non-intentional actions (Papaeliou & Trevarthen, 2006; Sakkalou & Gattis, 2012), and the results we present in Chapter 4 are the first to show that infants use prosody before words to communicate specific social emotions and

intentions like requests, statements, responses, and expressions of satisfaction or disapproval. Crucially, the fact that pragmatic uses of prosody develop before infants master the use of lexical cues offers two main conclusions: first, that prosody is the first grammatical component of language that infants use for communicative purposes; and second, that linguistic communication emerges before infants have the ability to use lexical items with semantic meanings.

Another important finding is that pre-lexical infants use prosody and gesture to interpret their interlocutors' intentionality. Previous research on action and intention understanding claimed that infants learn to understand the others' intentions through the information available in the shared action context. Our studies show that even in the absence of a disambiguating shared action context (a situation which is actually quite frequent in everyday interactions), prosody and gestures help infants to understand the intentional meaning of the others' actions. When they need to understand what an interlocutor means, infants certainly use contextual cues and common ground (Clark, 1996), but this is insufficient in many situations where shared action context is ambiguous or absent. Infants nonetheless regularly face such situations and have learned to understand what the other person means. We propose that young infants comprehend an interlocutor's intentionality not only through contextual cues but also by relating the form of the act-accompanying features such as prosody and gesture shape to their pragmatic function. These results are in accordance with usage-based accounts of Theory of Mind development (Liszkowski, 2013) which suggest that human interactions are central for infants

learning to predict the intentions of others. Other studies have claimed that mind-reading abilities are observed in young infants when these abilities are tested through tasks that do not require complex processing and explicit reasoning about a third-party's mental representations (Kovács et al., 2010; Kovács, 2009; Onishi & Baillargeon, 2005; Rakoczy, 2012).

Future research should investigate the relative contribution of prosody and gestures in early intentional communication. Our results showed that when these cues are combined, infants interpret them as conveying pragmatic meaning, but we still do not know the relative contribution of each of the two modalities. Previous studies suggested that multimodal combinations might be 'scaffolding' the comprehension of linguistic meaning that infants later learn to understand even if expressed only with speech means. This idea has been proposed in other studies investigating pre-school children's comprehension of more complex linguistic meanings such as evidentiality and epistemicity (Armstrong, Esteve-Gibert, & Prieto, 2014). Also, a great deal more research needs to be done on the early production of pragmatic aspects of prosody. We know quite a lot about how infants start producing pointing gestures and their importance in language and communicative development. But further research should address prosodic development using similar experimental paradigms to completely understand its role in language development.

In conclusion, this dissertation has contributed to a full understanding of the communicative uses of prosody and gesture in

early infancy. We have shown that the temporal integration of gesture and speech is present in infants from early stages of language and cognitive development, both in production and perception, and suggested that infants' development of prosodic abilities may trigger this temporal alignment. And, crucially, we have provided evidence that infants use prosody and gestures in an integrated way for intention understanding and to transmit their own social intentions, a reflex of an early stage in the development of language pragmatics.

7. REFERENCES

Alferink, I., & Gullberg, M. (2014). French-Dutch bilinguals do not maintain obligatory semantic distinctions: Evidence from placement verbs. *Bilingualism: Language and Cognition*, *17*, 22–37.

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing. *Language Resources and Evaluation*, ***41***, 3–4.

Armstrong, M., Esteve-Gibert, N., & Prieto, P. (2014). The acquisition of multimodal cues to disbelief. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the Speech Prosody 2014*. Dublin (Ireland).

Astruc, L., Payne, E., Post, B., Vanrell, M. M., & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech*, *56*(2), 229–253.

Aureli, T., Perucchini, P., & Genco, M. (2009). Children's understanding of communicative intentions in the middle of the second year of life. *Cognitive Development*, *24*(1), 1–12.

Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False belief understanding in infants. *Trends in Cognitive Sciences*, *14*, 110–118.

Balog, H. L., & Brentari, D. (2008). The relationship between early gestures and intonation. *First Language*, *28*(2), 141–163.

Balog, H. L., Roberts, F. & Snow, D. (2009). Discourse and intonation development in the first-word period. *Enfance, 3*, 293–304.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415–419.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "Theory of Mind"? *Cognition, 21*, 37–46.

Barth-Weingarten, D., Dehé, N., & Wichmann, A. (2009). *Where Prosody meets pragmatics*. Bingley: Emerald.

Bates, E. Benigni, L., Bretherton, I., Camaioni, L., & Volterra. V. (1979). *The emergence of symbols: Cognition and Communication in infancy*. New York: Academic Press.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, *8*(6), 492–499.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *The British Journal of Developmental Psychology*, *30*(3), 359–375.

Behrens, H., & Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language*, *32*(1), 1–34.

Bergmann, K., Aksu, V., & Kopp, S. (2011). The Relation of Speech and Gestures: Temporal Synchrony Follows Semantic Synchrony. In *Proceedings of the 2nd Workshop on Gesture and Speech in Interaction*. Bielefeld (Germany).

Blake, J. & Boysson-Bardies, B. De (1992). Patterns in babbling: a cross-linguistic study. *Journal of Child Language, 19*, 51–74.

Blake, J., O'Rourke, P., & Borzellino, G. (1994). Form and function in the development of pointing and reaching gestures. *Infant Behavior and Development*, *17*, 195–203.

Boersma, P., & Weenink, D. (2012). *Praat: doing phonetics by computer*. [http://www.praat.org/]

Bonsdroff, L., & Engstrand, O. (2005). Durational patterns produced by Swedish and American 18- and 24-month-olds: Implications for the acquisition of the quantity contrast. In *Papers from the 18th Swedish Phonetics Conference* (pp. 59–62). Gothenburg (Sweden).

Borràs-Comes, J., Kaland, C., Prieto, P., & Swerts, M. (2013). Audiovisual Correlates of Interrogativity: A Comparative Analysis of Catalan and Dutch. *Journal of Nonverbal Behavior*, *38*(1), 53–66.

Bosch, L., & Sebastian-Galles, N. (2001). Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy*, *2*, 29–49.

Boysson-Bardies, A. B. De, & Vihman, M. M. (1991). Adaptation to Language: Evidence from Babbling and First Words in Four Languages. *Language*, *67*(2), 297–319.

Brüne, M. (2005). Emotion recognition, "theory of mind", and social behavior in schizophrenia. *Psychiatry Research*, *133*, 135–147.

Bruner, J. S. (1975). The onthogenesis of speeh acts. *Journal of Child Language*, *2*(1), 1–19.

Bryant, G. H., & Barrett, H. C. (2007). Recognizing Intentions in Infant-Directed Speech: Evidence for Universals. *Psychological Science*, *18*(8), 746–775.

Buitelaar, J. K., & van der Wees, M. (1997). Are deficits in the decoding of affective cues and in mentalizing abilities independent? *Journal of Autism and Developmental Disorders*, *27*, 539–556.

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come toegher. In D. McNeill (Eds.), *Language and Gesture* (pp. 235–257). Chicago: Cambridge University Press.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*, 337–342.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. Campbell & G. T. Smith (Eds.), *Recent advances in the psychology of language: formal and experimental approaches* (pp. 347–360). New York: Plenum Press.

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The Role of Declarative Pointing in Developing a Theory of Mind. *Infancy*, *5*(3), 291–308.

Camaioni, L., Perucchini, P., Muratori, F., Parrini, B., & Cesari, A. (2003). The communicative use of pointing in autism: developmental profile and factors related to change. *European Psychiatry*, *18*(1), 6–12.

Capone, N. C., McGregor, K. K. (2004). Gesture development: A review for clinicians and researchers. *Journal of Speech Language and Hearing Research, 47*, 173–186.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2014). The relation between executive function and theory of mind is more than

skin deep. *Journal of Cognition and Development*. Published online. Doi: 10.1080/15248372.2013.824883.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*(2), 315–330.

Carpenter, M., & Liebal, K. (2011). Joint Attention, Communication, and Knowing Together in Infancy. In A. Seemann (Eds.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 159–181). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), 1–174.

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*(6), 379–383.

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain, & W. J. Barry (Eds.), P*roceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 315-320). Dudweiler: Pirrot.

Chen, L. M., & Kent, R. D. (2009). Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language, 36*(1), 73–84.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clements, A., & Perner, J. (1994). Understanding of Belief. *Cognitive Development, 9*, 377–395.

Cochet, H., & Vauclair, J. (2010). Features of spontaneous pointing gestures in toddlers. *Gesture, 10*(1), 86–107.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). *Habit 2000: A new program for testing infant perception and cognition*. (Version 2.2.5c). Austin: University of Texas.

Crespo-Sendra, V., Kaland, C., Swerts, M., & Prieto, P. (2013). Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics, 47*(1), 1–13.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language, 12*(3), 643–661.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language, 25*(2), 141–168.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language, 14*, 145–167.

D'Odorico, L., & Franco, F. (1991). Selective production of vocalization types in different communication contexts. *Journal of Child Language, 18*, 475–499.

Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic correlates of stress in babbling: an acoustical study. *Child Development, 71*(5), 1258–1270.

De Ruiter, J. P. (1998). *Gesture and speech production*. Doctoral Dissertation. Katholieke Universiteit, Nijmegen.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284–311). Cambridge University Press.

De Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate? *Advances in Speech-Language Pathology, 8*(2), 124–127.

DePaolis, R. a., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, *36*(2), 406–422.

Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy, *Developmental Science 4*(1), 40–48.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavioural categories: Origins, usage and coding. *Semiotica, 1*, 49–98.

Engstrand, O., & Bonsdroff, L. (2004). Quantity and duration in early speech: preliminary observations on three Swedish children. In *Papers from the 17th Swedish Phonetics Conference* (pp. 64–67). Stockholm (Sweden).

Engstrand, O., Williams, K., & Lacerda, F. (2003). Does Babbling Sound Native? Listener Responses to Vocalizations Produced by Swedish and American 12- and 18-Month-Olds. *Phonetica*, *60*, 17–44.

Escudero, D., Aguilar, L., Vanrell, M. M. & Prieto, P. (2012). Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system. *Speech Communication*, *54*(4), 566–582.

Esteve-Gibert, N., Liszkowski, U., & Prieto, P. (in press). Prosodic and gesture features distinguish the pragmatic meanings of pointing gestures in child-directed communication. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Approaches to intonational grammar in Ibero-Romance*. Amsterdam, NL: John Benjamins.

Esteve-Gibert, N., Pons, F., Bosch, L., & Prieto, P. (2014). Are gesture and prosodic prominences always coordinated?

Evidence from perception and production. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the Speech Prosody Conference* (pp. 222–226). Dublin (Ireland).

Esteve-Gibert, N., & Prieto, P. (2012). Esteve-Prieto Catalan corpus. [http://prosodia.upf.edu/phon/ca/corpora/description/estevepriето.html].

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements. *Journal of Speech, Language, and Hearing Research, 56*(3), 850–865.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: effects of contrast polarity. *PNAS*, *102*(47), 17245–17250.

Feldman, R., & Reznick, J. S. (1996). Maternal perception of infant intentionality at 4 and 8 months. *Infant Behavior and Development*, *19*, 483–496.

Fenson, F., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–173.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, *60*, 1497–1510.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Ferré, G. (2010). Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. In *Proceedings of the Language Resources and Evaluation. Workshop on Multimodal Corpora*, (pp. 86–91). Malta.

Flax, J., Lahey, M., Harris, K., & Boothroyd, A. (1991). Relations between prosodic variables and communicative functions. *Journal of Child Language 18*(1), 3–19.

Frota, S., & Vigário, M. (2008). The intonation of one-word and first two-word utterances in European Portuguese. In *Proceedings of the Third Conference on Tone and Intonation (TIE 3)*. Lisboa (Portugal).

Furrow, D. (1984). Yound children's use of prosody. *Journal of Child Language, 11*(1), 203–213.

Furrow, D., Podrouzek, W., & Moore, C. (1990). The acoustical analysis of children's use of prosody in assertive and directive contexts. *First Language, 10*(28), 37–49.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493–501.

Galligan, R. (1987). Intonation with single words: purposive and grammatical use. *Journal of Child Language, 14*(1), 1–21.

Goldman, A. I. (2009). Mirroring, Simulating and Mindreading. *Mind & Language, 24*(2), 235–252.

González-Fuente, S., Escandell-Vidal, V., & Prieto, P. (2014). Gestural codas lead to the interpretation of irony. In *Proccedings of From Sound to Gesture (S2G) Conference*. Padova (Italy).

Gopnik, A. (2003). The theory theory as an alternative to the innateness hypothesis. In L. Antony & N. Hornstein (Eds.), *Chomsky and his Critics*. New York: Basil Blackwell.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085–1108.

Gussenhoven, C. (2005). Transcription of Dutch Intonation. In Sun-Ah Jun (Eds.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Gussenhoven, C., Rietveld, T., Kerkhoff, J., & Terken, J. (2003). *ToDI, Transcription of Dutch Intonation* (second edition). [http://todi.let.kun.nl/ToDI/home.htm]

Haan, J., van Heuven, V. J., Pacilly, J., & van Bezooijen, R. (1997). An anatomy of Dutch question intonation. *Linguistics in the Netherlands*, *14*(1), 97–108.

Habets, B., Kita, S., Shao, Z., Ozyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854.

Halliday, M. A. K. (1975). *Learning how to mean: explorations in the development of language*. New York: Elsevier.

Harrison, S. (2010). Evidence for node and scope of negation on coverbal gesture. *Gesture*, *10*(1), 29–51.

Hauf, P. (2007). Infants perception and production of intentional actions. *Progress in Brain Research*, *164*, 285–301.

Henry, J. D., Phillips, L. H., Crawford, J. R., Ietswaart, M., & Summers, F. (2006). Theory of mind following traumatic brain injury: The role of emotion recognition and executive functions. *Neuropsychologia*, *44*, 1623–1628.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behavior & Development*, *32*(3), 262–274.

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*(3), 598–613.

Igualada, A., Bosch, L., & Prieto, P. (2014). Exploring the link between early multimodal communication abilities and vocabulary measures at 18 months of age. In *Proceedings of the Budapest CEU Conference on Cognitive Development*. Budapest (Hungary).

Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: precursor to the gesture-speech system? *Child Development*, *75*(4), 1053–1066.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371.

Iverson, J. M., Tencer, H. L., Lany, J., & Goldin-Meadow, S. (2000). The relation between gesture and speech in congenitally blind and sighted language-learners. *Journal of Nonverbal Behavior*, *24*, 105–130.

Iverson, J. M., & Thelen, E. (1999). Hand, Mouth and Brain. *Journal of Consciousness Studies*, *6*(11-12), 19–40.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19.

Jovanovic, B., & Schwarzer, G. (2007). Infant perception of the relative relevance of different manual actions. *European Journal of Developmental Psychology*, *4*(1), 111–125.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687.

Karousou, A. (2003). *Análisis de las vocalizaciones tempranas: su patrón evolutivo y su función determinante en la emergencia*

de la palabra. Doctoral Dissertation, Universidad Complutense de Madrid, Spain.

Kelly, S. D., Ozyurek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267.

Kendon, A. (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207–227). The Hague: Mouton.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kent, R. D., & Murray, D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, *72*(2), 353–365.

Keren-Portnoy, T., Majorano, M., & Vihman, M. M. (2009). From phonetics to phonology: the emergence of first words in Italian. *Journal of Child Language*, *36*(2), 235–267.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Eds.), *Language and Gesture* (pp. 162–185). Cambridge: Cambridge University Press.

Kita, S. (2003). *Pointing: where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.

Kita, S., & Özüyrek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language,* 48(1), 16–32.

Knudsen, B., & Liszkowski, U. (2012a). 18-Month-Olds Predict Specific Action Mistakes Through Attribution of False Belief, Not Ignorance, and Intervene Accordingly. *Infancy*, *17*(6), 672–691.

Knudsen, B., & Liszkowski, U. (2012b). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, *15*(1), 113–122.

Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese : The effect of infant-directed action on infants ' attention and object exploration. *Infant Behavior and Development*, *32*(4), 437–444.

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, *12*(1), 48–54.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*(3), 396–414.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us? In M. Zanna (Eds.), *Advances in experimental social psychology* (pp. 389–450). San Diego, CA: Academic Press.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, *41*(3), 841–849.

Leavens, D. (2009). Manual deixis in apes and humans. In C. Abry, A. Vilain, & J. L. Schwartz (Eds.), *Vocalize to localize* (pp. 67–86). Amsterdam: John Benjamins.

Leonard, T., & Cummins, F. (2010). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, *26*(10), 1457–1471.

Leung, E. H. L., & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology, 17*, 215–220.

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and Voicing in Deictic Expressions. *Journal of Memory and Language*, *24*, 133–164.

Levinson, S. C. (2006). On the human "interaction engine." In S. C. Levinson, & N. J. Enfield (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). Oxford: Bergs.

Levitt, A., & Utman, J. (1992). From babbling towards the sound systems of English and French: a longitudinal two-case study. *Journal of Child Language, 19*(1), 19–49.

Lewkowicz, D. J. (2010). Infants Perception of Audio-Visual Speech Synchrony. *Developmental Psychology*, *46*(1), 66–77.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science, 12*(2), 264–71.

Liebal, K., Carpenter, M., & Tomasello, M. (2011). Young children's understanding of markedness in non-verbal communication. *Journal of Child Language*, *38*(4), 888–903.

Liebal, K., & Tomasello, M. (2009). Infants appreciate the social intention behind a pointing gesture: Commentary on "Children's understanding of communicative intentions in the middle of the second year of life" by T. Aureli, P. Perucchini and M. Genco. *Cognitive Development, 24*(1), 13–15.

Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.

Liszkowski, U. (2005). Human twelve-month-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. *Gesture*, *5*(1), 135–154.

Liszkowski, U. (2007). Human twelve-months-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. In S. Pika (Eds.), *Gestural Communication in Nonhuman and Human Primates* (pp. 124–140). Amsterdam: John Benjamins.

Liszkowski, U. (2013). Using Theory of Mind. *Child Development Perspectives, 7*(2), 104–109.

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science, 7*(3), 297–307.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-Olds Point to Provide Information for Others. *Journal of Cognition and Development, 7*(2), 173–187.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out new news, old news, and absent referents at 12 months of age. *Developmental Science, 10*(2), F1–7.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition, 108*(3), 732–739.

Liszkowski, U., Schäfer, M., Carpenter, M., & Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science, 20*(5), 654–660.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology, 3*, 71–89.

MacWhinney, B., Pléh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology, 17*(2), 178–209.

Mahy, C. E. V, Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, *9C*, 68–81.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994–1997.

Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language*, *14*(2), 255–268.

Marslen-Wilson, W., & Tyler, L. K. (980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

Masataka, N. (2003). From index-finger extension to index-finger pointing: ontogenesis of pointing in preverbal infants. In S. Kita (Eds.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 69–84). Mahwah: Lawrence Erlbaum.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

Mehler, J., Jusczyk, P., & Lambertz, G. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*, 702–709.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining Facial Imitation: A Theoretical Model. *Early Development & Parenting*, *6*(3-4), 179–192.

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, *47*(6), 1028–1039.

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-Month-Olds Know What "We" Have Shared in a Special Way. *Infancy*, *13*(1), 90–101.

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, *43*(2), 309–317.

Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as a function of attentional focus. *Journal of Cognition and Development*, *2*, 109–129.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for Referential Understanding in the Emotions Domain at Twelve and Eighteen Months. *Child Development*, *72*(3), 718–735.

Moses, L. J., & Tahiroglu, D. (2010). Clarifying the relation between executive function and children's theories of mind. In B. Sokol, U. Muller, J. Carpendale, A. Young, & G. Iarocci (Eds.), *Self- and Social Regulation: Exploring the Relations between Social Interaction, Social Cognition, and the Development of Executive Functions* (pp. 218–231). Oxford: Oxford University Press.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*(3), 938–954.

Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274.

Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination: Longitudinal changes and predictive value on early lexical development. *Gesture, 12*(1), 16–39.

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communication. *Journal of Child Language*, *13*, 15–29.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance, 24*(3), 756–766.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language Discrimination by English-Learning 5-Month-Olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language*, *43*(1), 1–19.

Nobe, S. (1996). *Representational gestures, cognitive rhythms, and acoustic aspects of speech: a network/threshold model of gesture production*. Doctoral Dissertation. University of Chicago.

Oller, D. K., Wieman, L. A., Doyle, J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language, 3*, 1–11.

Olson, S. L., Bates, J. E., Bayles, K. (1982). Predicting long-term developmental outcomes from maternal perceptions of infant and toddler behavior. *Infant Behavior and Development, 12*, 77–92.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Ozçalişkan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101–113.

Ozyurek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and

gesture: insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616.

Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing "communication" and "apprehension." *Journal of Child Language*, *33*(1), 163–178.

Papaeliou, C., Minadakis, G., & Cavouras, D. (2002). Acoustic patterns of infant vocalizations expressing emotions and communicative functions. *Journal of Speech, Language, and Hearing Research*, *45*(2), 311–317.

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733.

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. M. (2011). Measuring Child Rhythm. *Language and Speech*, *55*(2), 203–229.

Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. Neuchâtel: Delachaux et Niestlé.

Piaget, J. (1953). *The Origin of Intelligence in Children*. New York: International University Press.

Pierrehumbert, J. (1980). *The Phonetics and Phonology of English Intonation*. Doctoral Dissertation. Massachusetts Institute of Technology.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication*. Cambridge, USA: MIT Press.

Pons, F., & Bosch, L. (2010). Stress pattern preference in Spanish-learning infants: the role of syllable weight. *Infancy*, *15*(3), 223–245.

Pons, F., & Lewkowicz, D. J. (2014). Infant perception of audio-visual speech synchrony in familiar and unfamiliar fluent speech. *Acta Psychologica, 149*, 142–147.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515–526.

Prieto, P. (2006). The relevance of metrical information in early prosodic word acquisition: a comparison of Catalan and Spanish. *Language and Speech, 49*(2), 233–261.

Prieto, P., Borràs-Comes, J., Cabré, T., Crespo-Sendra, V., Mascaró, I., Roseano, P., Sichel-Bazin, R., & Vanrell, M. M. (2013). Intonational phonology of Catalan and its dialectal varieties. In S. Frota, & P. Prieto (Eds.), *Intonational variation in Romance*. Oxford: Oxford University Press.

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language, 39*(2), 221–257.

Rakoczy, H. (2012). Do infants have a theory of mind? *The British Journal of Developmental Psychology, 30*(1), 59–74.

Ramus, F., Hauser, M. D., Miller, C. T., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and cotton-top tamarin monkeys. *Science, 288*(5465), 349–351.

Reddy, V., Markova, G., & Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PloS One, 8*(6), e65289.

Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: infants selectively respond to indirect emotional signals. *Child Development, 78*(2), 503–521.

Rochat, P. (2007). Intentional action arises from early reciprocal exchanges. *Acta Psychologica, 124*(1), 8–25.

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. (2008). The Speech Focus Position Effect on Jaw-Finger Coordination in a Pointing Task. *Journal of Speech Language and Hearing Research*, *51*(6), 1507-1521.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Warehem, T. (2006). Introducing Phon: a software solution for the study of phonological acquisition. In D. Bamman, T. Magnitskaia & Colleen Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Roustan, B., & Dohen, M. (2010). Gesture and Speech Coordination : The Influence of the Relationship Between Manual Gesture and Speech. *Proceedings of: INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*. Makuhari (Japan).

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, *12*(1), 182–187.

Rusiewicz, H. L. (2010). *The Role of Prosodic Stress and Speech Perturbation on the Temporal Synchronization of Speech and Deictic Gestures.* Doctoral Dissertation, University of Pittsburgh.

Russell, J. (1997). How executive disorders can bring about an inadequate 'theory of mind'. In J. Russell (Eds.), *Autism as an Executive Disorder* (pp. 256–304). New York: Oxford University Press.

Sachs, J. (1993). The emergence of intentional communication. In J. Gleason (Eds.), *The development of language*. New York: Macmillan.

Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, *27*(1), 1–16.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*, 3–11.

Sansavini, B., Guarini, S., & Stefanini, C. (2010). Early development of gestures, object-related actions, word comprehension and word production, and their relationships in Italian infants. *Gesture*, *10*(1), 52–85.

Sarriá, E. (1991). Observación de la comunicación intencional preverbal: un sistema de codificación basado en el concepto de la categoría natural. *Psicotema*, *3*, 359–380.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, *99*, 143–165.

Searle, J. (1976). A classification of illocutionary acts. *Language in Society*, *5*, 1–23.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671.

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2001). The origins of face perception: specific versus non-specific mechanisms. *Infant and Child Development*, *10*, 59–65.

Skoruppa, K., Pons, F., Bosch, L., Christophe, A., Cabrol, D., & Peperkamp, S. (2013). The Development of Word Stress Processing in French and Spanish Infants. *Language Learning and Development*, *9*(1), 88–104.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919.

Snow, D. (2006). Regression and Reorganization of Intonation Between 6 and 23 Months. *Child Development*, *77*(2), 281–296.

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, *112*(12), 1025–1058.

So, W. C., Demir, O. E., & Goldin-Meadow, S. (2010). When speech is ambiguous gesture steps in: Sensitivity to discourse-pragmatic principles in early childhood. *Applied Psycholinguistics*, *31*(1), 209–224.

Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, *12*(6), 1013–1019.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907–912.

Southgate, V., van Maanen, C., & Csibra, G. (2007). Infant pointing: communication to cooperate or communication to learn? *Child Development*, *78*(3), 735–740.

Sperry, L. A., & Symons, F. J. (2003). Maternal Judgments of Intentionality in Young Children with Autism: The Effects of Diagnostic Information and Stereotyped Behavior. *Journal of Autism and Developmental Disordorders*, *33*(3), 281–287.

Striano, T., Henning, A., & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies*, *7*, 251–271.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition* (pp. 217–262). San Diego, CA: Academic Press.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. MA: Harvard University Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463.

Toro, J. M., Trobalon, J. B., & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, *6*(2), 131–136.

Trainor, L. J., Austin, C. M., & Desjardins, N. (2000). Is Infant-Directed Speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195.

Trevarthen, C. (1977). Descriptive analyses of infant communicative behaviour. In H. R. Schaffer (Eds.), *Studies in mother-infant interaction*. London: Academic Press.

Trevarthen, C. (1979). Communication and cooperation in early infancy. A description of primary intersubjectivity. In M. Bullowa (Eds.), *Before speech: The beginning of human communication* (pp. 321–347). London: Cambridge University Press.

Trevarthen, C. (1982). The primary motives for cooperative understanding. In G. Butterworth, & P. Light (Eds.), *Social cognition: studies of the development of understanding.* Brighton: Harvester Press.

Trevarthen, C. (1990). Signs before speech. In T. A. Sebeok, & J.U. Sebeok (Eds.), *The semiotic web*. Berlin: Mouton de Gruyter.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Eds.), *Action, gesture and symbol: The emergence of language*. New York: Academic Press.

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science 7*(3), 261–269.

Vihman, M. M., & DePaolis, R. A. (1998). Perception and production in early vocal development: evidence from the acquisition of accent. In M. C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *Chicago Linguistic Society, 34*, 373–386.

Vihman, M. M., DePaolis, R. A., & Davis, B. L. (1998). Is There a Trochaic Bias in Early Word Learning? Evidence from Infant Producion in English and French. *Child Development*, *69*(4), 935–949.

Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2009). A dynamic systems approach to babbling and words. In E. L. Bavin (Eds.), *The Cambridge Handbook of Child Language* (pp. 163–182). Cambridge: Cambridge University Press.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., Simmons, H., & Miller, J. I. M. (1985). From Babbling to Speech : A Re-Assessment of the Continuity Issue. *Language*, *61*(2), 397–445.

Vihman, M. M., Nakai, S., & DePaolis, R. (2006). Getting the rhythm right : A cross-linguistic study of segmental duration in babbling and first words. *Laboratory Phonology*, *8*, 341–366.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual Language Discrimination in Infancy. *Science, 316*(5828), 1159.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*(3), 655–684.

West, B., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: a practical guide using statistical software*. New York: Chapman & Hall/CRC.

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*, *18*, 501–516.

Willems, R. M., Ozyurek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, *47*(4), 1992–2004.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraircing function of wrong bekfs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, *22*(2), 145–160.

Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. (2004). The timing of speech-accompanying gestures with respect to prosody. In *Proceedings of From Sound to Sense* (pp. 97–102). Cambridge, MA: Massachusetts Institute of Technology.

Appendix 1 (Introducció en català)

1. INTRODUCCIÓ

1.1. El desenvolupament de la comunicació intencional

Poc després de néixer els infants ja mostren comportaments socials primerencs, com ara el fet de preferir mirar cares humanes (Cassia, Turati, & Simion, 2004; Farroni et al., 2005; Johnson, Dziurawiec, Ellis, & Morton, 1991; Simion, Macchi Cassia, Turati, & Valenza, 2001), d'imitar les expressions facials dels adults (Meltzoff & Moore, 1983; Meltzoff & Moore, 1997), o de formar part de protoconverses amb adults en què hi ha alternança de torns (Levinson, 2006; Murray & Trevarthen, 1986; Striano, Henning, & Stahl, 2006). No obstant això, en aquesta mena d'interaccions els infants encara no consideren l'interlocutor com a un agent intencional.

Els investigadors estan d'acord en què és a partir dels 9 o 12 mesos d'edat que els infants comencen a comunicar-se intencionalment i a interpretar els actes dels altres com a intencionals (p. ex., Bates, Benigni, & Bretherton, 1979; Piaget, 1953). Segons Tomasello (1993), sobretot hi ha dos fets que indiquen l'inici de la comunicació intencional: 1) que els infants tinguin capacitat per a distingir el fins dels mitjans en les seves pròpies produccions i en

les dels altres, i 2) que els infants participin en actes d'acció conjunta.

El desenvolupament de la capacitat per a distingir els fins dels mitjans en les produccions pròpies i en les dels altres implica que els infants han d'aprendre a diferenciar l'acció (el mitjà) de la intenció que motiva aquesta acció (el fi). Els estudis sobre producció i percepció de les accions han vist que és cap als 6-9 mesos d'edat que els infants comencen a entendre les accions com a dirigides cap a un objectiu i que poden distingir entre accions intencionals i accions accidentals (p. ex., Carpenter, Akhtar, & Tomasello, 1998; Jovanovic & Schwarzer, 2007; Meltzoff, 1995; Woodward, 1998, 1999) (vegeu-ne una revisió completa a Hauf, 2007). Woodward (1999) va estudiar infants de 5 i 9 mesos en quatre condicions diferents per a veure si a aquesta edat podien distingir les accions amb un objectiu específic de les accions sense un objectiu específic. En dues de les condicions, l'infant o bé veia un actor que agafava un objecte de manera intencional (condició "agafar") o bé veia l'actor que tocava un objecte de manera accidental amb el revers de la mà (condició "revers de la mà"). La diferència entre els assajos de prova (*test trials*) i els de familiarització era que en els primers o bé s'ensenyava un objecte nou (condició "objecte nou") o bé es feien uns moviments nous per a tocar l'objecte (condició "nous moviments"). Després d'analitzar els temps de mirada, van veure que els infants de 9 mesos reaccionaven diferent segons si era una acció intencional o una d'accidental, i que preferien la condició en què es canviava d'objecte més que no pas la condició en què es canviava el

moviment per a agafar-lo. A més, també van veure que els infants de 5 mesos mostraven un patró similar però no tan clar, senyal de què a aquesta edat la capacitat per a detectar la intencionalitat de les accions comença a desenvolupar-se però encara no ho està totalment. Carpenter, Akhtar i Tomasello (1998) també van estudiar el desenvolupament de la capacitat per a distingir accions intencional de no intencionals per a veure si els infants es basaven en els comportament vocal que acompanyava l'acció. Carpenter et al. (1998) van mostrar que els infants de 16 mesos preferien imitar les accions intencionals, quan la única diferència entre les dues era que les intencionals anaven acompanyades del mot "*There!*" ("Allà!"), mentre que les accidentals del mot "*Whoops!*" ("Ui!"), i que cada mot anava acompanyat de les característiques prosòdiques corresponents.

El desenvolupament de la capacitat de participar en actes d'acció conjunta és un altre pas clau en el desenvolupament de l'infant com a agent intencional (p. ex. Bruner, 1975; Carpenter & Liebal, 2011; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Mundy & Newell, 2007; Mundy et al., 2007; Repacholi & Meltzoff, 2007; Senju & Csibra, 2008; Tomasello & Farrar, 1986; Tomasello, 1995). Un dels primers estudis experimentals que van investigar el sorgiment de les accions conjuntes va ser Trebarthen i Hubley (1978). Aquests autors van analitzar el comportament d'un infant durant el primer any de vida mentre interactuava amb un cuidador al laboratori. Els resultats van mostrar que abans dels 6 mesos l'infant interactuava només amb l'objecte o només amb l'adult, però mai alternava somriures o mirades entre els dos. Així, el

comportament de l'infant en aquesta etapa estava caracteritzat per interaccions diàdiques en què l'infant no veia l'altre com a un agent intencional. A partir dels 6 i fins als 9 mesos, l'infant començava a alternar l'atenció entre els dos elements, mirant la mare o somrient-li mentre manipulava l'objecte. I a partir dels 10 mesos, l'infant ja alternava l'atenció entre l'objecte i l'adult o somreia l'adult fent referència a l'objecte de manera habitual. Aquests resultats van mostrar que a aquesta edat l'infant ja establia interaccions triàdiques.

La manera com els infants desenvolupen l'habilitat per a entendre la intencionalitat s'ha estudiat abastament en la literatura sobre teoria de la ment. La teoria de la ment fa referència a l'habilitat per a inferir els estats mentals (és a dir, les intencions, els desitjos i les creences) d'un mateix i dels altres per tal de predir-ne el comportament (Premarck & Woodruff, 1978, i molts altres després). Sabem que als 9 mesos els infants distingeixen les accions amb un objectiu específic de les accions que no tenen un objectiu específic (Woodward, 1998, 1999). Els infants, però, també han d'aprendre a detectar les creences i els desitjos que motiven les accions intencionals dels altres per tal de "llegir la ment" dels altres. Tradicionalment s'han utilitzat tasques de falses creences (*false-belief tasks*), com ara la tasca Sally-Ann (Bahron-Cohen, Leslie, & Firth, 1986; Wimmer & Perner, 1983), per tal d'avaluar el desenvolupament infantil de la teoria de la ment. En aquesta tasca, els infants observen una escena en què un agent col·loca un objecte en un recipient i després marxa. Mentre aquest agent no hi és, un segon agent mou l'objecte cap a un altre recipient. Llavors el primer

agent torna a escena i l'investigador pregunta a l'infant on és que aquest agent buscarà l'objecte. Si l'infant assenyala el recipient on el primer agent ha deixat l'objecte abans de marxar vol dir que passa la tasca satisfactòriament, però si assenyala el recipient on el segon agent ha col·locat l'objecte, vol dir que no passa la tasca satisfactòriament. Els estudis que han utilitzat aquesta tasca mostren que els infants de 3 anys miren el recipient correcte però que no són capaços d'assenyalar-lo explícitament fins que no tenen 4 o 5 anys (p. ex. Clements & Perner, 1994; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983).

Per a passar aquesta tasca satisfactòriament cal que l'infant hagi adquirit habilitats mentals complexes, com ara processos d'inhibició (de la pròpia perspectiva per a generar-ne una altra mentre es manté la perspectiva rellevant a la memòria de treball) (Carlson, Claxon, & Moses, 2014; Russell, 1997), representacions neurals compartides (accedir directament als estats psicològics d'un mateix i dels altres i que això es reflecteixi en sistemes neuronals) (Gallese & Goldman, 1998; Goldman, 2009), o teorització sobre les relacions entre estats i accions (formar conceptes abstractes sobre estats mentals i accions) (Gopnik & Wellman, 2012; Gopnik, 2003; vegeu Mahy, Moses, & Pfeifer, 2014 per a una revisió completa dels tres punts). Arran d'aquesta complexitat, diversos investigadors han proposat que les tasques de falses creences no siguin l'única manera de mesurar les habilitats de teoria de la ment (Carlson et al., 2014; Moses & Tahiroglu, 2010). Llegir la ment de l'altre també implica inferir cognitivament les emocions de l'altre. Diversos estudis han mostrat que les tasques de detecció

d'emocions es correlacionen directament amb la comprensió de les intencions, i que totes dues habilitats es processen a les mateixes àrees cerebrals (Brüne, 2005; Buitelaar & van der Wees, 1997; Henry, Phillips, Crawford, Ietswaart, & Summers, 2006; Mier et al., 2010).

Alguns investigadors han creat tasques que requereixen menys esforç cognitiu per a investigar si els infants de menys de 3 anys ja tenen teoria de la ment (Baillargeon, Scott, & He, 2010; Buttelmann, Carpenter & Tomasello, 2009; Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005). En conjunt, aquests estudis mostren que els infants de menys de 3 o 4 anys ja tenen habilitats per inferir els estats mentals dels altres si se'ls avalua a partir de tasques més senzilles cognitivament. Onishi i Baillargeon (2005) van avaluar infants de 18 mesos en un experiment de violació d'expectatives. En aquest experiment es familiaritzava els infants en una tasca en què un agent amagava una joguina en una capsa. Aleshores, la joguina o bé es movia a una altra capsa mentre l'agent no hi era o bé es movia a l'altra capsa quan l'agent hi era present però tan aviat com l'agent marxava d'escena l'objecte tornava a la seva posició inicial. Després de mesurar els temps de mirada de l'infant, els resultats van indicar que els infants miraven més estona si l'agent buscava l'objecte allà on no se suposava que l'agent havia de buscar-lo en base a allò que l'agent havia pogut observar durant l'esdeveniment. Per tant, els autors van demostrar que amb només 18 mesos els infants ja fan prediccions de les creences dels altres.

D'entre els estudis que estan d'acord amb un desenvolupament primerenc de la teoria de la ment, hi ha dues perspectives diferents (tot i que relacionades): la innatista i la teoria basada en l'ús. La perspectiva innatista de la teoria de la ment proposa que els humans naixem amb una habilitat innata per a veure els altres com a agent socials amb intencions (p. ex., Kovács et al., 2010; Onishi & Baillargeon, 2005). Kovács et al. (2010), per exemple, van investigar si els infants de 7 mesos computaven automàticament l'atenció d'un agent respecte d'un objecte. En una sèrie d'estudis, els autors van observar que les reaccions implícites dels infants estaven influïdes per l'atenció de l'agent respecte dels objectes, fins i tot si la presència de l'agent era irrellevant en el context de l'acció. Van mostrar que els infants miraven més estona les accions si la seva pròpia expectativa d'un esdeveniment futur no es confirmava i, encara més important per a l'estudi, també si l'expectativa de l'agent sobre l'esdeveniment futur no es confirmava. Aquests resultats van indicar que els infants fan computacions automàtiques sobre les creences i tenen un "sentit social".

La teoria basada en l'ús en relació amb el desenvolupament de la teoria de la ment proposa que des d'etapes molt primerenques els infants aprenen que els altres són agents intencionals gràcies al fet que tenen experiències d'interacció social, i això els fa desenvolupar el sistema per a predir les accions. En aquest sentit, hi ha diversos estudis que mostren que els infants tenen expectatives flexibles sobre el comportament dels altres, i que aquestes expectatives depenen de la situació sociocontextual en què ocorren els comportaments. Específicament, mostren que els infants

reaccionen de manera flexible als comportaments dels altres (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Liszkowski, Carpenter, & Tomasello, 2008; Moll & Tomasello, 2007; Liszkowski, 2013; Southgate, Chevallier, & Csibra, 2010), que inicien accions de manera flexible segons els comportaments dels altres (Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004; Liszkowski, Carpenter, & Tomasello, 2007; Liszkowski, Schäfer, Carpenter, & Tomasello, 2009), i que intervenen activament en la modificació dels comportaments dels altres (Knudsen & Liszkowski, 2012a, 2012b).

En conclusió, la recerca sobre desenvolupament de la comunicació intencional indica que a partir de la segona meitat del primer any de vida els infants comencen a transmetre intencions als altres i a inferir les creences, desitjos i intencions dels altres. Tots aquests estudis mostren que la informació sociocontextual que precedeix l'acte intencional és un element bàsic perquè l'infant pugui atribuir intencionalitat a les seves pròpies accions i a les dels altres. Això no obstant, gairebé no s'ha investigat si els infants utilitzen les marques lingüístiques i gestuals que acompanyen les accions per a comunicar-se intencionalment i aquesta tesi vol aportar evidències científiques sobre aquest fet.

1.2. La prosòdia en el desenvolupament de la comunicació intencional

Tal com mostren els estudis de percepció, els infants són sensibles als trets prosòdics des de molt aviat en el desenvolupament del llenguatge. Se sap que els infants prefereixen les propietats prosòdiques de la parla dirigida als infants (p. ex. Fernald & Kuhl, 1987; Fernald, 1985). Fernald (1985) va observar que els infants de només 4 mesos d'edat preferien escoltar la parla dirigida als infants més que no pas la parla dirigida als adults, quan la única diferència entre les dues era les propietats prosòdiques, i Fernald i Kuhl (1987) van mirar si això era conseqüència dels patrons de freqüència fonamental (F0, més amplitud tonal), de durada (més durada) i d'amplitud (menys variabilitat) de la parla dirigida a infants. En aquest estudi es va veure que els infants prefereixen la parla dirigida a infants per les propietats prosòdiques que té: específicament, prefereixen els valors més elevats d'amplitud tonal, i no tant per les característiques de durada o amplitud.

També, els infants de només 3 mesos distingeixen dos llengües si formen part de categories rítmiques diferents (p. ex. Mehler, Jusczyk, & Lambertz, 1988; Nazzi, Bertoncini, & Mehler, 1998), una habilitat que és compartida amb altres espècies animals (p. ex. Ramus, Hauser, Miller, Morris, & Mehler, 2000; Toro, Trobalon, & Sebastián-Gallés, 2003). I als 4-5 mesos ja saben utilitzar els trets segmentals per a distingir llengües de la mateixa categoria rítmica. A més a més, des de ben petits són sensibles a la posició de la prominència prosòdica, i als 6-9 mesos d'edat ja prefereixen el

patró accentual de la seva llengua materna (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk, Cutler, & Redanz, 1993; Pons & Bosch, 2010).

Pel que fa a la producció de patrons prosòdics, sembla que segons quina dimensió prosòdica es tingui en compte (ritme, accent i entonació), el procés de desenvolupament ocorre a etapes diferents. Mampe, Friederici, Christophe i Wermke (2009) van investigar els patrons d'entonació i intensitat dels plors de nadons francesos i alemanys (mitjana d'edat de 3 dies) per a veure si diferien en funció de la llengua que havien escoltat abans de néixer. Els resultats d'aquest estudi van mostrar que, tal com s'esperava, tots els plors seguien un patró en forma d'arc, però que el pic de l'arc de la intensitat i la freqüència fonamental canviava significativament en funció de la llengua a la qual els infants havien exposats abans de néixer: el pic d'intensitat i entonació en els infants francesos era cap al final de l'arc (en forma de contorn ascendent), mentre que en els infants alemanys el pic d'intensitat i entonació era cap al començament de l'arc (en forma de contorn descendent).

Els patrons accentuals es desenvolupen una mica després en el procés d'adquisició del llenguatge (p. ex. Behrens & Gut, 2005; Davis, MacNeilage, Matyear, & Powell, 2000; DePaolis, Vihman, & Kunnari, 2008; Keren-Portnoy, Majorano, & Vihman, 2009; Snow, 2006; Vihman, DePaolis, & Davis, 1998; Vihman, Nakai, & DePaolis, 2006). Per exemple, Davis et al. (2000) van analitzar els paràmetres acústics de l'accent en el balboteig dels infants i van veure que, tot i que eren capaços de fer ús de la freqüència

fonamental, la intensitat i la durada per a marcar prominència, en aquesta etapa els infants encara no produïen els patrons acústics de prominència tal com ho fem els adults. Vihman et al. (1998) va comparar infants francesos i anglesos en el moment en què produïen 25 paraules (és a dir, a l'etapa de producció de les primeres paraules) i van veure que els infants francesos produïen més iambes (CV'CV) que troqueus ('CVCV), tal com fan els adults, mentre que els infants anglesos produïen tots dos patrons. Els autors de l'estudi proposaven una explicació: en aquesta etapa els infants ja estan influïts pel patró de la llengua del seu entorn, perquè en francès totes les paraules tenen l'accent a l'última síl·laba, mentre que en anglès tots dos patrons són presents a la llengua però hi ha preferència per l'accent a la penúltima síl·laba. Això no obstant, sembla que no és fins força més tard que els infants adquireixen els patrons rítmics de la seva llengua materna. Payne, Post, Astruc, Prieto i Vanrell (2011) van comparar els patrons rítmics d'infants de 2, 3 i 4 anys en català, castellà i anglès. L'anàlisi va mostrar que als 2 anys els infants semblaven utilitzar alguns trets rítmics d'acord amb la seva llengua materna (sobretot pel que fa a la variabilitat entre intervals), tot i que els resultats van millorar significativament a mesura que es van analitzar les edats posteriors.

La prosòdia, però, no és només una propietat acústica. Una de les funcions principals de la prosòdia és contribuir al significat pragmàtic de la oració. La prosòdia s'utilitza per a expressar l'actitud del parlant respecte d'un objecte o esdeveniment, per a distingir entre tipus oracionals, per a estructurar la informació a la frase, per a organitzar i mantenir interaccions, o per a transmetre

informació epistèmica i evidencial (per a una revisió completa sobre la interfície entre prosòdia i pragmàtica, vegeu Barth-Weingarten, Dehé, & Wichmann, 2009). Hi ha força estudis sobre el desenvolupament de la interfície entre prosòdia i pragmàtica que mostren que els infants no adquireixen un inventari complex de contorns entonatius amb una funció pragmàtica consistent i tal com fem els adults fins que no estan a l'etapa de producció de dues paraules (Chen & Fikkert, 2007; Frota & Vigário, 2008; Prieto, Estrella, Thorson, & Vanrell, 2012). Altres estudis sobre comprensió primerenca del significat pragmàtic de la prosòdia han vist que la comprensió dels trets prosòdics a nivell de frase s'adquireix bastant tard en el procés de desenvolupament del llenguatge (p. ex. Cruttenden, 1985; Cutler & Swinney, 1987; MacWhinney, Pléh, & Bates, 1985). De fet, Cutler i Swinney (1987) proposaven una "paradoxa performativa" per la qual els infants aprenen a emprar alguns contorns prosòdics de manera adequada segons la intenció pragmàtica abans que no pas aprenen a entendre els significats pragmàtics que transmeten aquests mateixos contorns.

Tanmateix, però, hi ha motius per pensar que els infants poden utilitzar la sensibilitat primerenca que tenen per a percebre patrons acústics de la prosòdia per a comprendre significats intencionals i per a comunicar-se amb els altres. Hi ha dos motius per a pensar-ho. Primer, el fet que hi ha estudis previs que mostren que els infants comprenen i produeixen accions intencionals a partir de la segona meitat del primer any de vida (Woodward, 1998, 1999). Segon, el fet que els infants poden emprar patrons prosòdics al cap de poc

mesos d'haver nascut, no només per a la percepció de trets acústics (Bosch & Sebastian-Galles, 2001; Fernald & Kuhl, 1987; Jusczyk et al., 1993; Mehler et al., 1988; Nazzi et al., 1998; Ramus et al., 2000), sinó també pel que fa a la producció de parla sense un significat concret (Mampe et al., 2009).

De fet, sembla que els infants empren la prosòdia com a eina per a la comunicació intencional (D'Odorico & Franco, 1991; Papaeliou, Minadakis, & Cavouras, 2002; Papaeliou & Trevarthen, 2006; Sakkalou & Gattis, 2012). En un estudi de comprensió, Sakkalou i Gattis (2012) van examinar si els infants de 14 i 18 mesos imitaven més les accions intencionals que les accidentals, basant-se en Carpenter et al. (1998). En aquesta ocasió, però, les autores van fer que les dues accions només es distingissin per les característiques prosòdiques de la paraula que acompanyava l'acció. En un primer experiment, les accions intencionals estaven acompanyades de la paraula "*There*" ("Allà"), molta amplitud i durada, mentre que les accions accidental estaven acompanyades de la paraula "*Whoops!*" ("Ui!"), poca amplitud i durada. En un segon experiment, van utilitzar la mateixa metodologia però sense informació lèxica que distingís les condicions. En tots dos experiments, les autores van observar que a totes dues edats els infants imitaven més les accions intencionals que les no intencionals, i que hi havia un efecte d'edat quan desapareixia la informació lèxica, ja que llavors els infants de 18 mesos tenien millors resultats que els de 14. Aquests results mostren que els infants entenen que la prosòdia té significat pragmàtic i poden relacionar-ho amb la intencionalitat als 14 mesos.

Papaeliou i Trevarthen (2006) van centrar-se a investigar el desenvolupament de la producció de patrons prosòdics en relació amb la intencionalitat. Específicament, volien veure si els paràmetres acústics de les vocalitzacions (la durada, els valors de freqüència fonamental del començament, final, màxim, mínim i mitjana de la vocalització, i el rang de desviació estàndard de la freqüència fonamental) reflectien el valor intencional o no intencional de les vocalitzacions. Van enregistrar quatre infants anglesos de 10 mesos en dues situacions diferents: mentre jugaven amb la mare i mentre jugaven tot sols. Van considerar que les vocalitzacions eren intencionals si, quan es produïen, els infants dirigien la mirada cap a la mare, produïen gestos comunicatius, seguien la mirada o el gest d'assenyalament de la mare, o eren conseqüència de comportar-se tal com la mare els demanava. Per contra, van considerar que les vocalitzacions no eren intencionals si es produïen mentre els infants aguantaven un objecte, inspeccionaven un objecte, o completaven un comportament precedent. Els resultats van confirmar la seva hipòtesi, ja que les vocalitzacions intencionals eren més curtes i amb majors valors de rang tonal que no pas les vocalitzacions no intencionals.

Malgrat aquests dos estudis que exploren el paper de la prosòdia en la producció i comprensió primerenques de significats pragmàtic, encara cal molta recerca per a entendre com es desenvolupa la relació entre prosòdia i pragmàtica en els infants. Els estudis previs mostren que els infants distingeixen les accions intencionals de les no intencionals a través de la producció i processament dels patrons prosòdics. Ara bé, encara no se sap si els infants utilitzen la

prosòdia per a transmetre i interpretar significats pragmàtics més específics quan interactuen socialment.

1.3. La gestualitat en el desenvolupament de la comunicació intencional

Una evidència clara de què l'infant esdevé un agent intencional és la seva capacitat de producció i comprensió del gest d'assenyalar. El gest d'assenyalar es refereix a l'extensió simultània de braç i dit índex cap a un objecte o esdeveniment, amb la final de dirigir l'atenció o el comportament d'una persona cap a aquest objecte o esdeveniment. Juntament amb el gest per aconseguir un objecte (o *reaching gesture*, en anglès, que es refereix a allargar el braç i obrir la mà cap a una entitat per tal de dirigir l'atenció del cuidador cap a aquesta entitat), formen el grup de gestos díctics. El gest d'assenyalar es considera una senyal clara de comunicació intencional perquè implica que qui el produeix té l'objectiu de redirigir l'atenció d'un interlocutor i que qui l'ha d'entendre ha de reconèixer que l'altre té l'objectiu de redirigir-li l'atenció (Bates, Camaioni, & Volterra, 1975; Kita, 2003; McNeill, 1992).

S'han identificat tres intencions socials[8] diferents que motiven el fet d'assenyalar cap a un objecte o un esdeveniment. Bates et al. (1975)

[8] Al llarg de la tesi, alternarem els termes "intenció social", "significat pragmàtic" i "intenció comunicativa" indistintament per a referir-nos al significat d'un acte comunicatiu que es produeix en un context d'interacció social i que està dirigir a un interlocutor concret.

van distingir entre el gest d'assenyalar declaratiu i l'imperatiu. La intenció declarativa s'empra per fer que l'interlocutor es fixi en una entitat externa, és a dir, s'utilitza l'entitat externa com a eina per a captar l'atenció de l'altre. La intenció imperativa ocorre quan s'empra l'altre per tal d'aconseguir un objecte, és a dir, s'utilitza l'interlocutor com a eina per a aconseguir l'objecte. A partir de la proposta de Bates et al. (1975), Tomasello, Carpenter i Liszkowski (2007) van proposar que el gest d'assenyalar era un acte comunicatiu que s'utilitza per a dirigir l'atenció d'un interlocutor cap a un objecte amb tres intencions socials concretes: 1) aportar informació a l'interlocutor que li pot ser rellevant (intenció "declarativa informativa"); 2) compartir l'atenció cap a un objecte o esdeveniment amb l'interlocutor (intenció "declarativa expressiva"), o bé 3) demanar un objecte a l'interlocutor (intenció "imperativa"). Imagineu-vos que l'Anna i en Jaume estan parlant. L'Anna està de cara una finestra i en Jaume d'esquena. Mentre parlen comença a ploure i l'Anna assenyala la finestra per fer que en Jaume s'hi fixi i així aportar-li una informació que li pot ser útil. Això seria un exemple d'intenció declarativa informativa. Ara imagineu-vos que vosaltres i jo estem mirant bocabadats un castell de focs. Jo assenyalo cap al cel perquè vull compartir amb vosaltres l'emoció de veure una figura que han dibuixat els petards al cel. Això seria un exemple d'intenció declarativa expressiva. I ara penseu en una situació en què estem dinant amb la meva família. Ma mare assenyala l'ampolla de vi perquè vol que li passi però ho fa sense dir-me res perquè està mastegant el menjar. Això seria un exemple d'intenció imperativa.

Diversos estudis han investigat la capacitat dels infants d'entendre la intenció social del gest d'assenyalar i han mostrat que ho poden fer si el context social de l'acció els dóna prou informació per a fer-ho (Aureli, Perucchini, & Genco, 2009; Behne, Carpenter, & Tomasello, 2005; Behne, Liszkowski, Carpenter, & Tomasello, 2012; Camaioni, Perucchini, Bellagamba, & Colonnesi, 2004). Camaioni et al. (2004) van observar que els infants es comportaven diferent si el gest d'assenyalar que se'ls dirigia tenia intenció imperativa o declarativa expressiva, quan la única diferència entre els dos era la informació sociocontextual que havia precedit el gest. Els autors també van trobar un efecte d'edat: els infants entenien la intenció imperativa als 12 mesos però no era fins als 15 mesos d'edat que entenien la intenció declarativa expressiva. Tanmateix, però, a Behne et al. (2012) es va comprovar que als 12 mesos els infants sí que entenen la intenció declarativa del gest d'assenyalar quan indica la localització d'un objecte que estava amagat, a més de demostrar que hi havia una correlació entre comprensió i producció.

Els infants comencen a produir gestos d'assenyalar cap als 10 o 12 mesos, tant amb intenció declarativa com imperativa. Cochet i Vauclair (2010) van estudiar a fons molts aspectes del desenvolupament del gest d'assenyalar en infants de 15 a 30 mesos a partir de tres tasques que permetien observar gestos d'assenyalar amb intenció imperativa, informativa i expressiva. Els autors van trobar tres resultats principals: 1) que els infants produïen els gestos d'assenyalar amb intenció expressiva i informativa acompanyats de vocalitzacions més que no pas en el cas dels gestos amb intenció imperativa, d'acord amb estudis anteriors que suggereixen que els

gestos amb intenció declarativa estan més relacionats amb el desenvolupament del llenguatge que no pas els que tenen intenció imperativa (Camaioni, Perucchini, Muratori, Parrini, & Cesari, 2003; Camaioni, et al., 2004); 2) que la durada dels gestos d'assenyalar era major quan tenien intenció declarativa que quan tenien intenció imperativa, cosa que els autors interpretaven com a conseqüència per haver de mantenir la interacció en el cas de la situació declarativa (tot i que els autors reconeixen que aquest resultat podia estar influït per altres variables de la interacció); i 3) que la forma de la mà indicava la intenció del gest d'assenyalar, ja que en el gest declaratiu s'allargava el dit índex mentre que en el gest imperatiu s'allargava la mà oberta.

En un seguit d'estudis, Liszkowski et al. (2004, 2006) van investigar si els infants de 12 mesos podien assenyalar amb una intenció social imperativa i expressiva en cas que el context social fos l'adequat. A Liszkowski et al. (2004) van dissenyar un experiment en què un adult reaccionava de manera diferent als gestos d'assenyalar d'un infant: compartint l'atenció conjunta, mirant només la cara de l'infant, mirant només l'esdeveniment assenyalat, o ignorant tant l'infant com l'esdeveniment. Així, van mostrar que als 12 mesos els infants assenyalen amb la finalitat de compartir l'interès amb l'adult sobre un objecte, ja que assenyalaven amb més freqüència si l'adult compartia l'atenció amb l'infant en comparació amb les altres condicions. I a Liszkowski et al. (2006) van investigar si infants de 12 i 18 mesos eren capaços d'assenyalar per a informar un adult sobre la ubicació d'un objecte

que havia caigut. Els autors van confirmar la seva hipòtesi, ja que van veure que els infants podien fer-ho a ambdues edats.

Cal destacar que la producció primerenca del gest d'assenyalar s'ha correlacionat positivament amb les habilitats lingüístiques i gramaticals posteriors de l'infant. Igualada, Bosch i Prieto (2014), per exemple, van confirmar i ampliar els resultats de Murillo i Belinchón (2012): els infants que als 12 mesos produïen més gestos d'assenyalar combinats amb vocalitzacions eren els que als 18 mesos tenien un major desenvolupament gramatical. Iverson i Goldin-Meadow (2005), juntament amb Ozçalişkan i Goldin-Meadow (2005) van analitzar els infants a l'època de transició entre una i dues paraules, i van veure que hi havia un tipus concret de combinació entre gest i parla que predeia el desenvolupament gramatical de l'infant: les combinacions en què el gest aportava informació suplementària respecte de la parla. També, Rowe i Goldin-Meadow (2009) van veure que la quantitat de combinacions gest-parla als 18 mesos predeia la complexitat sintàctica de les frases als 42 mesos d'edat.

1.4. La integració de prosòdia i gest en els adults

Com a parlants adults, la majoria dels gestos comunicatius no es produeixen de manera aïllada sinó en combinació amb parla. La parla i el gest fan un front comú des d'un punt de vista fonològic i pragmàtic per a transmetre el significat intencional en les interaccions humanes (p. ex. Kendon, 1980; McNeill, 1992). D'una

banda, l'alineació fonològica entre gest i parla fa referència al fet que la part més prominent del gest (ja sigui el moment àlgid –o *stroke*, en anglès– o l'àpex, vegeu la Figura 1)[9] coincideix amb la part més prominent de la parla, que normalment és la síl·laba tònica o accentuada de l'enunciat que acompanya el gest d'assenyalar (p. ex., De Ruiter, 2000; Levelt, Richardson, & La Heij, 1985; Loehr, 2012; Nobe, 1996; Rochet-Capellan, Laboissière, Galván, & Schwartz, 2008; Rusiewicz, 2010; Yasinnik, Renwick, & Shattuck-Hufnagel, 2004). Loehr (2012) va analitzar interaccions espontànies de parlants anglesos per a veure com alineaven gest i prosòdia a nivells diferents. A partir de McNeill (1992), els nivells que va tenir en compte per a l'anàlisi gestual van ser els àpexs, les fases gestuals, les frases gestuals i les unitats gestuals. L'anàlisi prosòdia es va fer seguint el model mètric autosegmental (Pierrehumbert, 1980), així que els nivells que es van tenir en compte van ser els accents tonals, les frases intermèdies i les frases entonatives. Els resultats de l'estudi van mostrar que els àpexs del gest predeien de forma fiable la presència d'un accent tonal, i que les frases gestuals es correlacionaven amb les frases intermèdies.

[9] Segons McNeill (2005), el moment àlgid d'un gest (anomenat *stroke* en anglès) típicament es correspon amb l'interval de més esforç gestual aparent, i l'"esforç" es determina segons alguns paràmetres com ara la força relativa del moviment o la tensió aparent de la forma de la mà. En el cas del gest d'assenyalar, per exemple, el moment àlgid del gest és l'interval de temps en què el braç està tan estès com és possible, mentre que l'àpex és punt concret dins el moment àlgid del gest en què el dit també està tan estès com és possible.

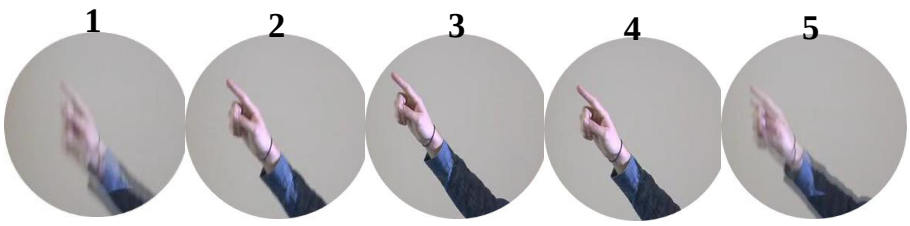

Figura 1. Fases d'un gest d'assenyalament: 1) preparació; 2-4) moment àlgid o *stroke*; 3) àpex; 5) retracció.

Les estructures prosòdiques i gestuals s'influeixen mútuament en tasques de percepció i producció. Krahmer i Swerts (2007) van investigar com influïa en la percepció i producció de prominència prosòdica el fet d'observar la prominència d'un gest rítmic (*beat gesture* en anglès)[10]. Els autors van veure que si es produïa un gest rítmic, això alterava la realització acústica de la prominència prosòdica, i que si els participants de l'experiment veien un gest rítmic, creien que la prominència prosòdica d'un estímul acústic que hi estava associat era major. La influència de l'estructura prosòdica en l'estructura gestual es va mostrar a Esteve-Gibert i Prieto (2013) en una tasca en què els participants assenyalaven a un punt mentre produïen una paraula concreta en un context de focus contrastiu. Les autores van veure que la posició de l'àpex del gest depenia de la posició del pic entonatiu dins la síl·laba accentuada, que alhora

[10] Un gest rítmic és un moviment bifàsic amb les mans o el cap que no té un significat semàntic concret però que acompanya la parla.

estava determinat per la presència d'una frontera entonativa abans o després de la síl·laba accentuada.

D'altra banda, l'alineació pragmàtica del gest i la parla es refereix al fet que el significat que transmet el gest és paral·lel o complementari al significat que transmet la parla. Kelly, Ozyurek i Maris (2010) van investigar la multimodalitat en la comprensió de conceptes per part de parlants en un context experimental. En l'experiment els participants observaven diverses accions representades amb gest i parla en tres condicions diferents: en una el gest i la parla transmetien informació congruent (parla: "esmicolar"; gest: esmicolar), en l'altra el gest i la parla transmetien informació totalment incongruent (parla: "esmicolar"; gest: tallar), i en la darrera les dues modalitats transmetien informació lleugerament incongruent (parla: "esmicolar"; gest: guerxar). Aquest estudi va mostrar que la comprensió de les accions està influïda pel nivell d'incongruència entre gest i parla, ja que com major era la incongruència menor era la comprensió dels conceptes, i que la influència del gest en la parla és inevitable. Així mateix, diversos estudis de neuroimatge han confirmat aquests resultats d'integració semanticopragmàtica entre gest i parla, i han mostrat que el cervell reaccions de manera diferent segons si el gest coincideix o no coincideix amb el contingut semàntic de l'ítem lèxic (Habets, Kita, Shao, Ozyurek, & Hagoort, 2011; Ozyurek, Willems, Kita, & Hagoort, 2007; Willems, Ozyurek, & Hagoort, 2009).

Hi ha un component concret de la parla, la prosòdia, que complementa o afegeix informació al significat del gest. Crespo-

Sendra, Kaland, Swerts i Prieto (2013) van estudiar l'ús de trets prosòdics i de gestos facials en parlants catalans i holandesos a l'hora de produir dos tipus d'oracions: oracions interrogatives totals i oracions interrogatives de contraexpectativa. Els autors van veure que els parlants de català utilitzaven més gestos facials que els holandesos per a marcar el tipus d'oració, i van relacionar aquest resultat amb el fet que en holandès s'empren més estratègies entonatives per a distingir entre tipus oracionals. D'aquesta manera, sembla que si una llengua no té trets prosòdics clars per a distingir entre tipus oracionals, es poden emprar estratègies gestuals per a ressaltar aquesta distinció. En un altre estudi es van obtenir resultats molt similars: a Borràs-Comes, Kaland, Prieto i Swerts (2013). Els autors van comparar oracions interrogatives totals amb declaratives informatives, també en holandès i català. En fer les comparacions els autors van observar que els gestos facials que acompanyaven la parla ajudaven els parlants de totes dues llengües, però els parlants d'holandès basaven els seus judicis en la informació acústica més que els parlants de català. Els autors van relacionar aquests resultats amb el fet que en holandès hi ha estratègies sintàctiques per a distingir els dos tipus d'oració analitzats.

El fet que gest i parla estiguin tan ben alineats des del punt de vista temporal, semàntic i pragmàtic ha fet que diversos investigadors proposin models de producció que ho intenten explicar. La teoria *Growth Point Theory*, que va proposar McNeill (1992), considera que el gest és un instrument comunicatiu i que tant gest com parla s'originen en el mateix imaginari mental i que de fet formen part d'un mateix sistema comunicatiu. El model *Sketch Model* (De

Ruiter, 2000) difereix del de McNeill perquè no considera que gest i parla vinguin del mateix imaginari mental. De Ruiter (2000) proposa que el gest i la parla es generen a partir de sistemes diferents però que interaccionen en una etapa molt inicial de la producció de la parla, just en el moment en què es planifica la intenció comunicativa. Segons aquest model, es planifica primer la realització del gest que no pas la parla, així que és el gest que influeix la parla i no a l'inrevés. Tanmateix, no tots els models consideren que el gest és un instrument comunicatiu. El model de Krauss, Chen i Chawla (1996) proposa que els gestos no són comunicatius i que la seva funció és la d'ajudar en el procés de l'accés al lèxic. Els autors diuen que la integració temporal de gest i prosòdia és resultat de l'estadi articulatori en què el codificador fonològic influeix els moviments motrius. Finalment, Kita i Ozyurek (2003) van exposar l'*Interface Model*, segons el qual gest i parla vénen de sistemes diferents, que aquests sistemes interactuen en moments diferents de la formulació del missatge, i que la codificació de la parla té un impacte en la formulació del gest.

1.5. La integració de prosòdia i gest en els infants

No se sap massa coses sobre el desenvolupament infantil de la integració fonològica i pragmàtica de gest i parla. Iverson i Thelen (2005) van proposar quatre fases que explicaven la dinàmica de la integració de gest i parla en els infants. Segons les autores, durant

els primers mesos de vida els infants mostren *lligams inicials* entre el sistema manual i l'oral. Això es veu clarament quan els infants s'acosten la mà a la zona de la cara o es fiquen els dits a la boca, quan obren la boca si se'ls pressiona el palmell de la mà (fenomen anomenat reflex Babkin), o també quan cap als 2 mesos es fiquen objectes a la boca. Després, cap als 3 o 4 mesos, i especialment a partir dels 6, els infants tenen un *control emergent* dels dos sistemes: comencen a produir moviments rítmics amb les mans i els braços que s'alineen amb balboteig. Més tard, als 10 o 11 mesos els infants mostren un *acoblament flexible* de gest i parla en què ja utilitzen les dues modalitats per a comunicar-se intencionalment. Les autores afirmen que aquesta etapa es caracteritza per una asimetria pel que fa al control i l'ús de les dues modalitats, ja que els infants empren els gestos més sovint que no pas la parla per a transmetre les seves intencions, i també pel fet que l'ús del gest prediu el desenvolupament lingüístic i gramatical posterior. Finalment, emergeix l'etapa d'*acoblament sincrònic*, en la qual els infants coordinen gest i parla per a finalitats comunicatives tal com ho fan els adults, és a dir, coordinant les prominències de les dues modalitats.

De fet, la integració audiovisual (AV) ja s'observa en etapes molt primerenques del desenvolupament cognitiu dels infants. Hi ha diversos estudis que han investigat aquesta qüestió a partir de gestos articulatoris acompanyats de parla. Aquests estudis han vist que infants de pocs mesos ja poden detectar asincronies temporals de 500 ms quan la informació visual precedeix la informació acústica (la parla) en gestos articulatoris (Lewkowicz, 2010; Pons &

Lewkowicz, 2014). Lewkowicz (2010) va mostrar que els nens de 4 a 10 mesos podien detectar asincronies AV d'una síl·laba articulada si el decalatge temporal era de 366 ms (i la parla precedia el gest), però això només si prèviament els nens havien estat exposats a sincronies majors. Després, Pons i Lewkowicz (2014) van veure que als infants catalans i espanyols de 8 mesos eren sensibles a l'asincronia AV en parla fluïda quan l'àudio precedia el vídeo en 366, 500 o 666 ms. També, van veure que aquest efecte era independent de la seva experiència lingüística anterior. Sembla, a més, que aquesta capacitat primerenca per a detectar asincronies és molt rellevant a l'hora de discriminar entre gestos articulatoris i contrastos fonètics, i a l'hora de segmentar la parla (Hollich, Newman & Jusczyk, 2005; Teinonen, Aslin, Alku & Csibra, 2008; Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastián-Gallés, & Werker, 2007).

La integració temporal de gest i parla en infants també s'ha començat a investigar des del punt de vista de la producció. Butcher i Goldin-Meadow (2000) van analitzar les produccions de gest i parlar de sis infants anglesos mentre interactuaven espontàniament amb un adult. Van trobar que al final del període de producció d'una paraula, els infants combinaven gest amb parla, però que en aquell moment encara no alineaven temporalment les prominències de les dues modalitats. Aquesta alineació temporal només apareixia quan els nens eren a l'etapa de producció de dues paraules. Aquest estudi ha sigut el primer i l'únic a investigar aquesta qüestió i, tot i els resultats interessants, no se'n poden extreure conclusions massa determinants perquè la base de dades no era homogènia i perquè per

a investigar l'alineació temporal no es van tenir en compte els estudis més recents sobre aquest tema. Altres estudis sobre el desenvolupament de les combinacions de gest i parla han recalcat la importància d'aquest tipus de produccions multimodals en el desenvolupament lingüístic i gramatical posterior dels infants (Igualada et al., 2014; Iverson & Goldin-Meadow, 2005; Murillo & Belinchón, 2012; Özçalişkan & Goldin-Meadow, 2005; Rowe & Goldin-Meadow, 2009).

Tot i els resultats que hem exposat en els paràgrafs anteriors, la literatura prèvia encara no explica del tot com es desenvolupa la integració del gest i la parla (i, sobretot, la prosòdia) per tal que els infants utilitzin les dues modalitats per a comunicar-se intencionalment. No hi ha cap estudi que hagi investigat la sensibilitat dels infants a l'alineació de gest i parla a partir d'estímuls comunicatius com ara un gest d'assenyalar acompanyat de parla en què l'alineació es caracteritzi per la coordinació de prominències. Així mateix, no hi ha estudis que se centrin en si els infants empren la integració de gest i parla per a produir i comprendre comunicació intencional en les etapes més primerenques del desenvolupament lingüístic i cognitiu. Se sap que als 10 mesos els infants utilitzen trets fonètics de la prosòdia per a marcar una vocalització com a intencional (Papaeliou & Trevarthen, 2006), que als 14 mesos utilitzen la prosòdia per a saber si una vocalització és intencional o no (Sakkalou & Gattis, 2012), i que entre els 14 i els 18 mesos entenen el significat intencional dels gestos comunicatius com el gest d'assenyalar si hi ha prou informació sociocontextual per a fer-ho (Aureli et al., 2009; Behne

et al., 2012; Camaioni et al., 2004; Liebal & Tomasello, 2009). Ara bé, el que no sabem és si els infants utilitzen el gest i la prosòdia de manera integrada per a transmetre i entendre les intencions socials dels altres, independentment de la informació sociocontextual.

1.6. Objectius generals, preguntes de recerca i hipòtesis

Aquest tesi pretén investigar la manera com els infants integren la prosòdia i el gest amb finalitats comunicatives abans de ser capaços d'emprar recursos lèxics. Específicament, ens interessa veure si els des de ben petits els infants integren temporalment els trets prosòdics i els gestos d'assenyalar des d'un punt de vista perceptiu i productiu, i si empren aquesta integració per a comunicar-se de forma intencional amb el seu entorn i per a entendre els actes intencionals que els són dirigits.

Ens centrarem en quatre preguntes de recerca, cada una de les quals es tractarà en un capítol diferent:

1) Els infants, des de ben petits, alineen temporalment gest i prosòdia en un context de comunicació intencional?

2) Són sensibles a l'alineació temporal entre la prominència prosòdica i la gestual?

3) Abans d'utilitzar ítems lèxics empren estratègies prosòdiques i gestuals per a marcar la intencionalitat de les vocalitzacions i per a expressar intencions socials?

4) Aquests infants també empren estratègies prosòdiques i gestuals de manera integrada per a entendre les intencions socials que se'ls dirigeixen?

Les nostres hipòtesis són les següents: a) que els infants poden integrar temporalment el gest i la parla des del moment en què comencen a produir combinacions de gest d'assenyalar amb vocalitzacions, b) que abans de produir combinacions de gest i parla ja són sensibles al fet que les dues modalitats han d'estar ben alineades temporalment, c) que els infants utilitzen aquesta integració multimodal per a transmetre intencions socials específiques, i d) que també utilitzen la integració multimodal per a la comprensió primerenca de la intencionalitat. La tesi s'organitza en quatre estudis independents, que es presenten del capítol 2 al 5. Els primers dos estudis (capítols 2 i 3) estudien la integració temporal primerenca de la prosòdia i el gest d'assenyalar en els infants. I els darrers dos estudis (capítols 4 i 5) estudien com els infants comencen a utilitzar la integració entre prosòdia i gest per a la comprensió i la producció de significats intencionals.

Específicament, el primer estudi (**capítol 2**) pretén investigar com els infants alineen gest i prosòdia a l'hora de comunicar-se amb els altres. La literatura anterior suggeria que aquesta alineació es produeix en el període de transició entre la producció d'una i de dues paraules, però en aquests estudis no s'havia analitzat

l'alineació temporal de manera gaire precisa i la mostra analitzada no era homogènia. Per a solucionar aquestes qüestions hem analitzat els trets prosòdics i gestuals en els actes intencionals de quatre nens mentre interactuaven espontàniament amb els seus pares, des del període del balboteig fins al final del període de producció d'una paraula (dels 11 als 19 mesos). L'anàlisi mostra tres resultats principals: a) els infants combinen gest i parla des del moment en què comencen a produir les primeres paraules; b) la majoria dels gestos que es combinen amb parla són gestos d'assenyalar amb intenció declarativa, i c) en aquestes combinacions de gest d'assenyalar i parla ja hi ha una alineació temporal molt semblant a la que fem els adults (és a dir, el gest comença abans que la parla, l'inici de la prominència gestual coincideix amb l'inici de la prominència prosòdica, i l'àpex del gest ocorre abans que acabi la prominència prosòdica).

Però si volem tenir una visió més completa de la integració temporal de gest i parla per part dels infants, també hem d'estudiar aquest tema des d'un punt de vista perceptiu. En el segon estudi (**capítol 3**) ens interessem per la sensibilitat dels infants a l'hora de detectar l'alineació temporal entre la prominència gestual i la prosòdica. Els estudis previs havien mostrat que els infants de 4 a 10 mesos podien percebre l'alineació temporal dels gestos articulatoris i la parla que els acompanya, però no s'havia mirat encara si en aquestes edats tan primerenques també són sensibles a l'alineació entre l'estructura prosòdica i la gestual en gestos comunicatius. Hem emprat el paradigma de *head-turn* amb infants de 9 mesos per a veure si a aquesta edat ja percebien la desalineació

entre prominències gestual i prosòdica, és a dir, si detectaven que el moment àlgid del gest no coincidia amb la síl·laba accentuada en gestos d'assenyalar acompanyats de parla. Els resultats de l'estudi mostren que amb només 9 mesos els infants ja són capaços de detectar l'alineació temporal entre les dues modalitats.

En el tercer estudi d'aquesta tesi (**capítol 4**) s'estudia si els infants que encara no produeixen les primeres paraules poden utilitzar estratègies prosòdiques (i gestuals) per a comunicar-se intencionalment. Fins ara altres estudis havien mostrat que els infants utilitzen la prosòdia per a diferenciar els actes de parla intencionals dels no intencionals, però no se sabia si en aquesta etapa també podien fer servir la prosòdia per a indicar la intenció pragmàtica específica que pretenien comunicar. Amb aquesta finalitat, vam enregistrar un corpus de quatre infants catalanoparlants mentre interactuaven espontàniament amb els seus pares a casa i es van analitzar les dades de tal manera que mostressin si els infants feien servir certs patrons prosòdics per a indicar si la vocalització era intencional i la intenció concreta que volien transmetre. Durant l'anàlisi es van tenir en compte aspectes acústics (de durada i amplitud tonal), gestuals i pragmàtics. Els resultats han mostrat que les vocalitzacions intencionals tenien patrons prosòdics diferents en comparació amb les vocalitzacions no intencionals, i que es podien distingir els significats pragmàtics específics dels actes de parla dels infants en base a les característiques prosòdiques de les vocalitzacions. Així, les peticions i expressions de descontentament tenien més amplitud tonal i més durada, mentre que les vocalitzacions declaratives i les

respostes tenien un rang tonal i durada menors. Aquest estudi és una de les primeres evidències de què la interfície entre prosòdia i pragmàtica es desenvolupa en etapes molt primerenques.

En el quart estudi (**capítol 5**) s'ha investigat si la capacitat per a emprar la prosòdia i el gest en la comunicació intencional primerenca també es pot observar pel que fa a la comprensió. Els estudis anteriors havien mostrat que els infants utilitzen la informació sociocontextual per a comprendre la intencionalitat de l'interlocutor, i el que nosaltres pretenïem era examinar si les propietats prosòdiques i gestuals dels actes de parla també podien ser rellevants per als infants a l'hora d'entendre les intencions dels altres. Per a fer-ho, es van dissenyar dos experiments amb infants de 12 mesos per a veure si podien distingir les intencions socials del gest d'assenyalar (expressiva, informativa o imperativa) a partir de les característiques prosòdiques i gestuals que acompanyaven el gest, tot controlant la informació sociocontextual que precedia l'acció (experiment 1) i també la informació lèxica (experiment 2). Els resultats han mostrat que els infants entenien la intenció de l'interlocutor, ja que reaccionaven d'acord amb la intenció subjacent de cada element díctic, i que la forma de la mà (gest d'assenyalar amb el palmell de la mà obert o amb el dit índex estirat) i les propietats prosòdiques (de durada i amplitud tonal) tenien un paper clau en aquest fet. Així, aquest estudi és el primer que mostra que els trets sociocontextual són importants però no indispensables, ja que els infants poden emprar la prosòdia i la forma del gest per a interpretar el significat intencional dels actes comunicatius.

Appendix 2 (Discussió i conclusions en català)

6. DISCUSSIÓ I CONCLUSIONS GENERALS

6.1. Resum dels resultats

Aquest tesi pretenia investigar com els infants integren la prosòdia i la gestualitat des del punt de vista temporal i com utilitzen aquesta integració per a la comunicació intencional. Hem presentat quatre estudis, cadascun dels quals en un capítol diferent. Els primers dos estudis s'han centrat en el desenvolupament de l'alineació temporal entre gest i parla en etapes primerenques de l'adquisició del llenguatge (capítols 2 i 3). I els darrers dos estudis s'han centrat en com els infants utilitzen les dues modalitats de manera integrada a l'hora de comunicar-se abans de ser capaços de produir els primers ítems lèxics (capítols 4 i 5).

Pel que fa el desenvolupament primerenc de l'alineació temporal de gest i parla, hem pogut observar dos resultats principals. Primer, en el capítol 2 hem vist que els infants produeixen gestos d'assenyalar en combinació amb vocalitzacions des dels primers mesos en què s'ha dut a terme l'anàlisi (és a dir, als 11 mesos), però que de fet la majoria gestos d'assenyalar apareixen en combinació amb parla a partir dels 15 mesos. També hem vist que els infants alineen temporalment prosòdia i gest de manera molt precisa des del moment en què produeixen les primeres combinacions de gest i parla, ja que el gest comença abans que la vocalització, l'inici del

moment àlgid del gest (l'*stroke*) s'alinea amb l'inici de la síl·laba accentuada, i els àpexs ocorren abans que acabin les síl·labes accentuades. Segon, en el capítol 3 hem vist que força abans de produir les primeres combinacions multimodals, els infants de 9 mesos ja són sensibles a l'alineació temporal que les caracteritza. En aquest sentit, hem vist que a aquesta edat ja poden distingir entre estímuls en què les prominències estan alineades correctament (és a dir, estímuls en què la prominència gestual coincideix amb la prominència prosòdica) d'estímuls que no estan correctament alineats (és a dir, estímuls en què la prominència gestual i la prosòdica no coincideixen).

Pel que fa a l'ús de la integració de gest i prosòdia amb finalitats comunicatives, també s'han pogut observar dos resultats principals. Primer, en el capítol 4 hem vist que els infants utilitzaven patrons prosòdics (sovint acompanyats de gestos comunicatius) abans que ítems lèxics per a marcar la intencionalitat de les vocalitzacions. Concretament hem trobat que el rang tonal i la durada de les vocalitzacions marcava si l'acte de parla era una petició, una resposta, una oració declarativa, una expressió de satisfacció o una de descontentament. Segon, en el capítol 5 hem mostrat que els infants poden emprar els trets prosòdics (entonació, rang tonal i durada) i de forma de la mà (gest d'assenyalar amb la mà oberta o amb el dit índex) per a inferir la intencionalitat d'una acció que els és dirigida, fins i tot si el context compartit que precedeix l'acció no els dóna informació prou rellevant. En conjunt, els resultats d'aquests dos capítols mostren que els infants no només utilitzen la informació contextual que precedeix una acció per a inferir-ne la

intencionalitat, sinó que també empren informació lingüística com ara la prosòdia o característiques comunicatives com ara la forma de la mà.

A les seccions següents posarem aquests resultats en relació amb la literatura prèvia sobre el tema i presentarem quina és la contribució que fan els nostres resultats a la recerca que s'havia fet fins ara: primer, pel que fa a la integració temporal de gest i prosòdia i, segon, pel que fa al desenvolupament de la comunicació intencional.

6.2. El desenvolupament de la integració temporal entre prosòdia i gest

La recerca sobre la integració de gest i parla proposa que el gest i la parla formen un sistema integrat en la comunicació humana (Kelly et al., 2010; Kendon, 1980; McNeill, 1992). Un dels motius pels quals s'ha fet aquesta afirmació és el fet que gest i parla estan alineats des del punt de vista temporal. En les converses espontànies, els gestos normalment apareixen en combinació amb parla i, quan es combinen, les respectives prominències solen estar alineades. Tot i que el concepte de prominència s'ha entès diferent segons l'estudi d'integració de gest i parla, sembla que la majoria d'investigadors coincideixen en què la síl·laba tònica o accentuada és el punt d'ancoratge clau a la parla (p. ex. De Ruiter, 2000; Levelt et al., Loehr, 2012; Nobe, 1996; Rochet-Capellan et al 2008; Rusiewicz, 2010; Yasinnik et al., 2003). En els gestos, normalment

la prominència s'ha entès com el moment àlgid del gest o *stroke* (l'interval de temps en què hi ha més esforç físic en el gest i que transmet més informació lingüística; Krahmer & Swerts, 2007; Rochet-Capellan et al., 2008), o l'àpex del gest (el punt, no l'interval, de més esforç en el gest; De Ruiter, 1998; Esteve-Gibert & Prieto, 2013; Loehr, 2012). Tot i la divergència de criteris, hi ha poc dubte de què en les combinacions de gest i parla la prominència gestual i la prosòdica estan temporalment alineades.

Els primers dos estudis que es presenten a la tesi mostren que l'alineació temporal de la prominència gestual i la prosòdica apareix de manera molt primerenca en el desenvolupament lingüístic i cognitiu. Específicament, aquests dos estudis mostren que el desenvolupament de l'alineació temporal multimodal es basa en el desenvolupament de l'estructura rítmica de la parla. En percepció s'ha vist que els infants de 9 mesos poden percebre el patró accentual de la seva llengua materna (Höhle et al., 2009; Jusczyk et al., 1993; Nazzi et al., 1998; Pons & Bosch, 2010; Skoruppa et al., 2009, 2013), i els nostres resultats mostren que a aquesta edat també són sensibles a l'alineació entre la prominència d'un gest d'assenyalar i la prominència prosòdica de la parla. En producció, se sap que és al període d'una paraula que els infants desenvolupen el patró accentual de la seva llengua materna (Behrens & Gut, 2005; Davis et al., 2000; Snow, 2006; Vihman et al., 1998), i els nostres resultats mostren que en aquest moment els infants també comencen a combinar la majoria dels gestos comunicatius que produeixen amb vocalitzacions, i que aquestes combinacions estan temporalment alineades. La hipòtesi que se'n deriva és que el desenvolupament de

l'estructura rítmica de la parla és determinant a l'hora que els infants hi puguin ancorar la prominència del gest.

Hi ha dues conclusions principals que s'infereixen d'aquests resultats. Primer, que el desenvolupament de l'alineació entre prosòdia i gest es relaciona amb el desenvolupament dels patrons prosòdics. Tan aviat com els infants són sensibles a la posició de la prominència prosòdia també són sensibles a la prominència gestual en gestos acompanyats de parla i al fet que les dues prominències han de coincidir. També, tan aviat com els infants empren marques acústiques de prominència per a produir els patrons accentuals de la seva llengua, també combinen els gestos comunicatius i la parla amb les corresponents prominències ben alineades. Segon, i essencial per a la recerca sobre integració de gest i parla, que el sistema gestual i el sistema de la parla estan integrats ja des d'etapes molt primerenques del desenvolupament lingüístic i cognitiu, com a mínim a partir de què els infants tenen 9 mesos. És per tot això que els resultats presentats en aquesta tesi apunten a què el gest i la parla són de fet un únic sistema en comunicació humana (De Ruiter, 2000; Kelly et al., 2010; Kita, 2003; McNeill, 2005).

En un futur la recerca amb altres tipus de gestos comunicatius i en estudis sobre la base neurocognitiva de l'alineació ens permetrà tenir una visió més completa sobre el desenvolupament de l'alineació temporal entre gest i parla. Pel que sabem, encara no s'ha investigat com els nens alineen temporalment els gestos icònics amb la parla corresponent quan els comencen a produir cap als 3 anys d'edat. També, encara no hi ha cap estudi que investigui

l'alineació temporal de gest i prosòdia en els gestos rítmics o *beats* o els moviments afirmatius del cap, que apareixen força més tard en el desenvolupament comunicatiu dels infants (cap als 4 o 5 anys). Finalment, encara cal avançar en la recerca sobre com l'alineació temporal interactua amb l'alineació semanticopragmàtica durant la infància. En estudis amb adults s'ha vist que hi ha certs gestos relacionats amb la parla en què les dues modalitats no estan alineades temporalment si la situació discursiva o les implicacions emotives ho requereixen (Bergmann et al., 2011; Esteve-Gibert et al., 2014; González-Fuente et al., 2014). Els nostres resultats, en combinació amb més recerca sobre altres tipus de gest i sobre les bases neurocognitives de l'alineació temporal, ens permetran saber si l'alineació temporal de gest i parla és només un fet motriu o perceptiu o si, a més d'això, també és una habilitat estretament lligada a la comprensió i producció del llenguatge.

6.3. La integració de prosòdia i gest en la comunicació intencional primerenca

Hi ha un consens general en el fet que tant prosòdia com gestualitat són elements de transmissió de significat intencional. Els parlants utilitzen la prosòdia per a identificar el tipus oracional (per exemple, per a marcar si és una oració interrogativa o declarativa), per a estructurar la informació a la parla (per exemple, per a distingir el tema del rema), per a expressar emocions, i per a comunicar significats com ara evidencialitat i epistemicitat. I els

parlants usen la gestualitat per a finalitats díctiques en relació amb un acte de parla (per exemple, a través d'un gest d'assenyalar amb intenció imperativa, expressiva o informativa), per a finalitats representacionals (a través de gestos icònics o metafòrics), o per a ajudar a organitzar la informació en el discurs (amb gestos rítmics o *beats*).

La literatura sobre desenvolupament cognitiu considera que la primera senyal de què l'infant s'està convertint en un agent intencional és l'habilitat que té per a produir i comprendre gestos d'assenyalar. D'una banda, els infants comencen a produir gestos d'assenyalar comunicatius entre els 8 i 12 mesos d'edat (Bates et al., 1979), i hi ha diversos estudis que mostren que als 12 mesos els infants ja poden produir gestos d'assenyalar amb diverses intencions socials: és a dir, amb la intenció de demanar un objecte a un adult (intenció imperativa), amb la intenció de compartir l'interès sobre un objecte amb un adult (intenció expressiva), o amb la intenció d'aportar informació rellevant per a l'adult (intenció informativa) (Liszkowski et al., 2004, 2006; Liszkowski, 2005). De l'altra, sembla que els infants de 12 a 14 mesos ja comprenen el significat intencional dels gestos d'assenyalar (Aureli et al., 2009; Behne et al., 2012; Camaioni et al., 2004). Aquests estudis mostren que els infants són capaços d'inferir la intenció dels gestos d'assenyalar comunicatius a través de la informació que els aporta el context social de l'acció que precedeix el gest d'assenyalar. A tall d'exemple, si l'infant està jugant amb un objecte i algú de cop i volta li aparta l'objecte de les mans, el gest d'assenyalar que possiblement produirà l'infant s'entendrà com a gest amb significat

imperatiu. Així mateix, si el context social indica que un adult necessita un objecte i llavors aquest adult assenyala cap a l'objecte, l'infant inferirà que aquest gest té un significat imperatiu.

Aquesta tesi ha mostrat que la prosòdia, i no només els gestos d'assenyalar, són un indicador de què l'infant s'està convertint en un agent intencional. Gairebé ja ningú dubta que la prosòdia és part de la gramàtica. La prosòdia identifica tipus oracionals, estructura la informació en el discurs i s'empra per a expressar emocions i significats pragmàtics. La recerca sobre desenvolupament dels significats pragmàtics de la prosòdia ha demostrat que des de ben petits els infants empren patrons prosòdics diferents segons si la vocalització acompanya una acció intencional o una acció accidental (Papaeliou & Trevarthen, 2006; Sakkalou & Gattis, 2012). Els resultats que hem mostrat en el capítol 4 són els primers a demostrar que els infants utilitzen la prosòdia abans que les paraules per a comunicar emocions i intencions socials determinades, com ara les peticions, les respostes, les oracions declaratives i les expressions de satisfacció o de descontentament. El fet que els infants utilitzin la prosòdia per a finalitats pragmàtiques abans que no pas utilitzin ítems lèxics ens porta a dues conclusions principals: primer, que la prosòdia és el primer component gramatical del llenguatge que els infants empren amb fins comunicatius; i segon, que la comunicació intencional emergeix abans que els infants tinguin la capacitat per a emprar ítems lèxics amb significat semàntic.

Un altre resultat que cal destacar és el fet que els infants que encara no produeixen les primeres paraules puguin emprar la prosòdia i la gestualitat per a interpretar la intencionalitat de l'interlocutor. Els estudis previs sobre comprensió d'accions i intencions havien afirmat que els infants entenen les intencions dels altres a través de la informació que poden extreure del context compartit de l'acció. Els estudis presentats en aquesta tesi mostren que si la informació que es pot extreure del context compartit no és prou informativa per a desfer l'ambigüitat de la intenció (cosa que és força habitual en les interaccions espontànies), els infants empren la prosòdia i la gestualitat per a entendre el significat intencional de les accions dels altres. A l'hora d'entendre quina és la intenció dels altres, segur que els infants utilitzen les marques contextuals compartides (Clark, 1996). Però aquestes marques sovint són insuficients en moments en què el context compartit de l'acció és ambigu o fins i tot absent. Tanmateix, els infants es troben amb aquesta mena de situacions i han d'aprendre a inferir les intencions dels altres. En aquesta tesi proposem que des de ben petits els infants entenen la intencionalitat dels seus interlocutors no només gràcies a les marques contextuals sinó també gràcies a aprendre a relacionar la forma de les característiques prosòdiques i gestuals que acompanyen els actes amb una funció pragmàtica. Aquestes observacions són coherents amb teories basades en l'ús sobre el desenvolupament de la teoria de la ment (Liszkowski, 2013). Aquest marc teòric proposa que les interaccions humanes són bàsiques per als infants a l'hora d'aprendre a predir les intencions dels altres. De fet, hi ha altres estudis que han observat que infants ben petits tenen habilitats

relacionades amb la teoria de la ment si aquestes habilitats s'avaluen a partir de tasques que no impliquen un processament complex ni un raonament explícit sobre les representacions mentals de terceres parts (Kovács et al., 2010; Kovács, 2009; Onishi & Baillargeon, 2005; Rakoczy, 2012).

En un futur caldria investigar la contribució relativa de la prosòdia i la gestualitat en el desenvolupament primerenc de la comunicació intencional. Els nostres resultats mostren que quan aquests dos elements estan combinats, els infants els atorguen un significat pragmàtic. Però encara no sabem quin és el paper exacte de la prosòdia i la gestualitat en aquesta interpretació de significat. En aquest sentit, estudis previs amb infants més grans (de 3, 4 i 5 anys) semblen indicar que el gest actua com a "suport" a l'hora de comprendre significats pragmàtics més complexos (com ara l'evidencialitat i l'epistemicitat) que més endavant els infants ja comprenen encara que només s'expressin a través de parla (Armstrong, Esteve-Gibert, & Prieto, 2014). També, cal fer més recerca sobre el desenvolupament primerenc del valor pragmàtic de la prosòdia. Se sap força bé com els infants comencen a produir gestos d'assenyalar i la importància que té aquest fet per al seu desenvolupament lingüístic i comunicatiu. Creiem que en un futur s'haurà d'estudiar el desenvolupament de la prosòdia a través de paradigmes experimentals similars que permetin tenir una visió més completa del paper que té aquest component de la gramàtica en el desenvolupament del llenguatge.

En conclusió, aquesta tesi és una contribució l'estudi del desenvolupament dels usos comunicatius de la prosòdia i la gestualitat. Hem mostrat que des de les etapes més inicials del desenvolupament lingüístic i comunicatiu els infants integren temporalment el gest i la parla tant pel que fa a la producció com pel que fa a la percepció, i hem indicat que el desenvolupament de les habilitats prosòdiques pot ser un factor decisiu en la integració temporal de les dues modalitats. Finalment, hem aportat evidències de què els infants empren prosòdia i gestualitat de manera integrada per a la comprensió de les intencions dels altres i per a la transmissió als altres de les seves pròpies intencions, cosa que reflecteix el desenvolupament primerenc de la pragmàtica lingüística.