Vocal-Affective Weaning and its role in musical enculturation: children's interest in music beyond the Infant-Directed Register

Juan Pablo Emilio Robledo del Canto

Wolfson College

September, 2019

This dissertation is submitted for the degree of Doctor of Philosophy

Preface

Declaration:

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

ABSTRACT

Well-established theoretical approaches to the Infant-directed Register (IDReg) draw a more or less tacit causal relation between it and all future musical engagement: the former establishes a basis for the latter. However, a mechanistic, observable chain of events linking these two phenomena has only been assumed, remaining for the most part unproblematized and undescribed. This thesis represents a first systematic attempt to fill this theoretical and empirical gap by concretely linking early musicality and the first manifestations of the most characteristic and widespread Western form of musical engagement: listening to recorded music. It does so by introducing a new construct as a mediating motivational factor— Vocal-Affective Weaning (VAW) —which mainly concerns the variation of caregivers' use of the IDReg across developmental time.

The thesis is grounded and tested in three literature reviews, two theoretical chapters, and three empirical ones. In terms of main findings, little evidence is found to support the existence of VAW as depicted in the main thesis. Consequently, any direct relationship between VAW and toddlers' attention to recorded music seems doubtful. Instead, data evidences a progressive use of Infant-Directed Speech (IDSp) as an ostensive cue used in the context of Natural Pedagogy, which allows for a better understanding of the relative importance of affectivity and cognition as parallel, coexisting governing principles that exert an influence on developmental changes concerning parental use of IDSp. If anything, data highlights interaction as a much more promising element in an explanatory chain linking the IDReg and Western forms of music engagement. The mentioned mismatch between the main thesis and results is interpreted in terms of the former's unjustified stress on 'centripetal' over 'centrifugal' attachment dynamics. Resulting data also allows to delineate more nuanced factorial and developmental accounts of toddlers' sustained attention to musical stimuli than has been previously advanced.

En memoria de mi abuela,
Sonia del Carmen.

Acknowledgments

There are so many people I am so grateful to. I think of you often and fondly, and it is an enormous pleasure mentioning you here.

I would like to thank my piano mentor back in Chile, professor Constanza Rosas and my conservatory mate Tomás, for making me a musician. As in my MPhil thesis, I would like to thank all of the LIF lab crew who patiently trained me as a psychologist, a researcher, and a lab member: David, Alejandra, Paloma, Javiera, Daniela, Himmbler, Camilo, Esteban and, of course, Carlos. I would like to thank the Cambridge Commonwealth, European & International Trust who unexpectedly and kindly funded most of my first year in the UK, and CONICYT for funding all the rest. To all of the people from the Centre for Music and Science— Becky, Erica, Xi, Mustafa, Sarah, Matthew, Mark, Satinder, Adam, David, Jiaxi, Li-Ching and Andrew —, thank you for so many seminars and dinners as well as for all of the sincere, respectful and thoughtful debate you taught me to be part of.

In getting the thesis done, acknowledgments to Danielle Daidone for her kind help in dealing with Praat scripts. Even more so to Dr David Greatrex, without whose expert advice on coding and statistics the second and third studies would not have been possible. My eternal gratitude to the late Sir Patrick Bateson for his invaluable impressions on the relationship between imprinting and attachment theories. Acknowledgments to Dr Dina Kronhaus for carefully and sincerely reviewing my second year report. To my fellow PhD students Becky, Ditlev and James, thank you for your company along the way. To Professor Elizabeth Tolbert, a huge thank you not just for the academic advice and thorough theoretical discussion, but also for your kind, kind support. An equally big thank you to my advisor, Professor Fabia Franco. Thank you for your constant interest and expert advice, your scepticism, and for the many coffees and delicious dinners talking about the PhD's status and possible directions.

I would like dearly to thank Wolfson College and all the people who welcomed me when I was just a foreign, disoriented fresher. Thank you Pascal, Uroš, Simone, Jenna, Maria Clara, Nina Marie, Misja, Camilla, Caspar, Josh, Tatiana and αδερφός μου for being there when I needed it. Thank you to Lyn Alcántara, James Downs, Andrei Şmid, Roger Briscoe and Linda Wicker, Conrad and Judith, Chris Lawrence, and Andrew Goldman for the so-far most enjoyed years of music-making in my life. Thank you! Thank you to the dear Chilean friends with whom we shared life abroad for many years: José Manuel and Camila, Marcos and Dana, Nico, Pablo, Megan and most of all, Elisa, Felipe and little Chimuelo.

Un grand merci aux Leblond, pas seulement pour leur soutien pendant le doctorat, mais surtout d'être devenus ma famille ici en Europe. A mi familia y amigos que me esperan siempre en Chile a través de los años de esta travesía que elegí emprender tan lejos, tantas, tantas gracias. Los siento cerca. Ustedes saben quiénes son, y cuánto los amo y extraño (‘*la mamá es la mamá!’*). Credit where it’s due, a warm thanks to my father Marcos who, long ago, seeded in me the precious gift of curiosity, along with the earnest ambition to satisfy it.

Finally, I would like to express my most sincere gratitude to professor Ian Cross, my supervisor. Ian has been the most calm and supportive mentor, and I feel deeply honoured to have worked with him.

Por último, a mi compañera en esta vida,

·—·

TABLE OF CONTENTS

Appendix F.
The Language Use Inventory for Young Children
Appendix G.
Focus group transcription
Appendix H.
Vocal data
Appendix I.
Musical stimuli
Appendix J.
Post-hoc interviews

INTRODUCTION

1. A long road and its map

Most of us probably cannot remember the first time we listened to music. It was simply always there, and it was always simply ‘music’. Similarly, in our daily adult lives, most people find listening to the music they like as easy and as pleasurable as riding a bicycle. It is, more often than not, simply a matter of pushing the ‘play’ button and enjoying. However, like in so many areas of human life, even the fact that we are able to enjoy recorded music represents the end of a long and forgotten road we all had to walk ourselves step by step. This thesis concerns, broadly speaking, this road— the road to ‘musical enculturation’.

Through musical enculturation, we learn to conceive of and recognize music as our significant others do. We learn what to listen to as well as when, where, and how to do it. The field of music psychology (the discipline of map makers that deals with this road, among many others) tells us that musical enculturation starts early in infancy: partly by means of innate capacities known as ‘musicality’, and partly by interacting with our caregivers. Regarding the latter, music psychology tells us that our first significant musical experiences correspond to paying attention (in general) and listening (in particular) to the special kind of vocalization through which caregivers spontaneously address non-verbal infants and children, known as the ‘Infant-Directed Register’ [IDReg].

The available map may thus tell us the location of the starting point of the road that musical enculturation represents. However, the situation quickly becomes complicated when we consider the rest of this road, as the map is full of blank spaces. Taking as an example the very first situation I described, we do not exactly know which factors play a role in taking us along the stretch between the start point of the road (the infant-directed register) and the point at which we are able, apparently without effort, to listen to and enjoy a record. Given that listening to recorded music happens to be the most common form of musical engagement in our Western society, we will focus on this ‘stretch of the road’ and the gap in knowledge in human development that it represents. This gap will become our research focus with the aim of arriving at a thesis which will, at least partially, fill it.

2. Navigating the thesis

Thus far, I have presented ideas in the simplest possible terms in order to provide an intuitive, initial frame of reference. However, in order to deal with the subject at an academic level, I will need to progressively add numerous further layers of complexity. Rather than properly laying out the subject in the introduction, I will properly do so in the first chapter. The remaining part of the introduction will, instead, consist of an outline of the seven chapters of the thesis and final discussion, and will give details of the structure and narrative style of the chapters and sections.

2.1. Thesis outline

As previously stated, I will need to deal with the subject by adding layers of increasing complexity in order to produce something that will actually be useful for other map makers. The first four chapters comprise such a task. Chapter 1 will properly introduce the subject by defining three essential notions: music, musicality and musical enculturation. We will see that, similar to the question of the chicken and the egg, music and musicality are interrelated in an ancient cycle in which they mutually lay the foundations for one another through the process of musical enculturation. I will then proceed to address a particular gap in knowledge and contemplate its consequences for the field of music psychology. I will make this gap in knowledge our subject, and this thesis will constitute an effort in filling in such a gap. This exercise will raise certain preliminary research questions, such as: what is the origin of children's motivation for listening to recorded, non-infant-directed Western music? Furthermore, could we pinpoint some kind of continuity between children's initial motivation to engage with the infant-directed register, and their later motivation for engaging with recorded Western music?

Our first approach to music, musicality and musical enculturation will place the infant-directed register at the centre of our research. Given that focussing on these components will compel us to rely heavily on this register as an articulating element, it will be necessary to reach an accordingly comprehensive understanding of it. By this token, the entirety of the technicalities and terminologies coined and accumulated by those who have studied the infant-directed register from different disciplines will constitute some of the aforementioned layers of complexity to be added. In particular, it will be necessary to understand the two dimensions of the register that are most relevant for our purposes—communication and attachment— before properly addressing infant-directed speech and scrutinizing its relationship with music. In order to allow for a firm and thorough understanding of these dimensions, I will proceed gradually through the second, third and fourth chapters, respectively. Although such a choice of pace will require more time, it will allow us to comfortably deal with the additional layers of complexity with as much clarity as possible.

Thus, chapter 2 will consist of a comprehensive review of animal and human communication. In the first part of this chapter, we will learn what constitutes animal signals and biological communication. In the second part, we will come to understand how human beings continue to employ such forms of communication on a daily basis through prosody, in parallel to 'word language'. This is crucial, taking into account that it is precisely through prosody that the infant-directed register's features and charm primarily manifest. Finally, I will discuss widespread literature that defends that the specifics of human vocal communication evolved largely as a response to the specifics of human caregiving.

The third chapter will comprise a second review dealing precisely with human caregiving and bonding. I will review John Bowlby's erection of attachment theory, as well as later criticisms and theoretical developments stemming from his initial scope. The chapter will also review some other attachment-related phenomena such as mentalisation, epistemic trust, and the bonding quality of interpersonal synchrony. In the second part of this chapter, I will approach attachment from a communicative point of view, relying on concepts from the previous chapter. The comparative analysis of thus far unconnected literature in this chapter will allow us to conceive of the infant-directed register as a form of biological communication at the service of attachment. Finally, this chapter will also introduce parent-offspring conflict, a construct that will prove essential to the formulation of the main thesis.

Relying on our conception of the infant-directed register firmly grounded in its evolution as an asset for communication and attachment, chapter 4 will begin by describing the register's features and functions, including its use as an ostensive cue as described in the theory of natural pedagogy (ToNP). Then, it will continue by critically addressing the exact association between the infant-directed register and music, and wonder whether it is the musical aspects of the former that lend it its characteristically high level of attractiveness. This chapter will reveal such a notion to be misleading, instead shifting the focus towards the register's characteristic 'loving tone' as the feature that seems to capture children's attention the most.

Chapter 5 constitutes the hinge of the present work. It is so because of representing a peak of complexity in the understanding of our problem, which is immediately followed by an exercise in simplification, necessary for generating research questions that can be empirically tested. In doing so, the chapter will lead us to the second part of this thesis that comprises three empirical studies (chapters 6, 7 and 8) as well as the final discussion. Such a shift towards the 'loving tone' and its unfolding through time will allow us to reformulate the research questions laid out in chapter 1. This will raise new questions, such as: when do infants start to consistently engage with music that is *not* harnessed into the IDReg— for example, recorded music?

In our main thesis, caregivers' variation in the use of IDSp will be explained as the result of the interplay between attachment and parent-offspring conflict in the context of the child's language acquisition and general enculturation. For such a variation that has so far only been tangentially approached in the literature, I will coin a name— 'vocal affective weaning' —and tentatively define it as

> the process of gradually introducing a human child to what will be its adult, culturally defined baseline intake of vocally expressed affection.

The reasoning behind this is that vocal-affective weaning seems to entail a form of continuity between early musicality and later tendencies to listen to recorded Western music, while at the same time explaining the motivation to do so. My main thesis is that recorded music— an instance culturally-sanctioned to be especially affective —should constitute a means for filling the vocal-affective gap that VAW could generate. Attachment inflicts on children a need for stable affection, while Western culture dictates that language should stand progressively less as a source of it. As a result of this, children will settle for alternatives to the IDReg in order to maintain balance in terms of their affective balance. One such culturally sanctioned alternative is recorded music, which is ubiquitous in our society.

The third and last part of chapter 5 deals with the operationalization of the thesis and the sketching of a research programme. Children's engagement with music will be operationalized in terms of cumulative fixation. Because I have postulated VAW as the main structuring element and the link between early musicality and recorded music, it will be VAW's evolution through time that we will focus on. The sparse, existing literature that informs this evolution points towards 'quantal changes' in parental prosody around the second year. As a consequence, the operationalized version of the thesis suggests that there may be quantal changes in caregivers' use of the IDReg (and its characteristic positive affect) between the ages of 18 and 24 months of age, the time around which children reach major linguistic, socio-linguistic, social and developmental milestones. The thesis also posits that there might be an increase in children's sustained attention to recorded music (music outside the IDReg) as a means for compensating such quantal changes. Before proceeding to the empirical work, the chapter will also provide a brief characterisation of what the literature tells us are toddlers' typical musical activities— even though our scope detaches from them. Finally, a methodological overview will be offered, providing rationale for and explaining the circumstances that led to the sequential implementation of the three following empirical studies.

Chapter 6 reports the first empirical study which was designed to test the applicability of the thesis, consisting of a focus group of ten caregivers. The aim of this study was to obtain the caregivers' perceptions, opinions, beliefs, and attitudes concerning their use of the IDReg, as well as assessing their own impressions on the idea of vocal-affective weaning. Participants' testimonies revealed little consciousness of any decrease in the vocal expression of affection as described in VAW. Instead, testimonies converged in depicting the role of mentalisation and the use of IDSp as an ostensive cue. Testimonies also consistently reported an increase in the expression of negative emotions such as anger as a direct consequence of acknowledging language acquisition. The above arguments will encourage us to reformulate VAW as the process of gradually introducing a human child to what will be its adult, culturally defined baseline intake of vocally expressed affection, as well as negative vocal-affective intake— particularly anger.

Chapter 7 details the findings of a second empirical study, this time consisting of a partial replication of Lamont's (2008) experience-sampling study which examines 18 to 24-month-old children's engagement with music. Results will suggest no relation between attention to music and children's language acquisition, the latter used as a proxy for VAW. Instead, results will reveal that the only variable that reliably predicts sustained attention in this age bracket is interaction with significant others— mainly the mother. The study also suggests a possible role of parent-offspring conflict, epistemic trust, and cultural learning in sustained attention to recorded music, as well as evidencing rather abundant quantities of infant-directed singing [IDSi] still being delivered by caregivers.

The last empirical study, reported in chapter 8, consists of six longitudinal case studies designed to directly assess children's sustained attention to music, pragmatic competence and parental prosody. Results show that participating girls around 23-months-old are already capable of paying sustained attention to recorded music without the mediation of interaction significant others. As in the case of the previous study, the data also suggests no connection to children's pragmatic development or parental prosody as described in the VAW hypothesis. Instead, results provide further support for the role of mentalisation, cultural learning in children's sustained attention to recorded music, as well as suggesting increments in the use of IDSp as an ostensive cue during this developmental window.

In the final discussion, I will aim to review and critically address the totality of this dissertation. Doing so will allow us to estimate which of its elements, if any, could be applied to current research in music psychology. I will first proceed by reviewing the main thesis one last time, selecting and examining the empirical findings that have alternatively supported or cast doubt on its different elements. The most important aspects of the postulated role of

VAW in musical enculturation will not be confirmed by the empirical studies designed to test it. Instead, the use of IDSp as a tool for encouraging toddlers' familiarisation with opaque cultural products— including recorded music —will be highlighted. The reasons behind such a mismatch between the main thesis and data will be discussed and related to my theoretical understanding of attachment theory. Empirical data will also provide a better overall picture of the development of children's engagement with recorded music. As a conclusion, I will continue to defend the relevance of the subject described and addressed in this dissertation, and will advocate the undertaking of more theoretical and empirical efforts to fill this particular blank space in the proverbial map.

2.2. A note on language and structure

Given that the present dissertation will deal with the interrelation and generation of rather complex ideas, a further strategy for facilitating reading is the use of an appropriate style of language. Always within the boundaries of academic writing, I have chosen to write in the simplest possible terms with no particular academic expertise taken for granted, thus allowing us to avoid the obscurity of field-dependent conventions and assumptions. Far from underestimating the reader, this style of writing follows the principle that the better understood a problem, the simpler its rendition may be. As a rule of thumb and in order to prompt the reader's involvement, contents will be narrated using an active voice in the first-person plural 'we'. The first-person singular will be reserved as a resource for drawing attention to my exclusive responsibility when it comes to choices which I have made as a researcher. An exception to this general style will be found in the methodological sections of the three chapters that correspond to empirical studies (chapters 5, 6 and 7), where a passive voice (traditionally associated to reports and papers) will be employed.

Once again, because we will often be dealing with the interrelation of contents sometimes scattered chapters apart, the location of these contents will be systematically indicated. Most chapters are subdivided into sections (in this very case, 2. Navigating the thesis), subsections (2.2. A note on language and structure), and so on. As an exception, chapters two and three, which are longer and more varied in content than the rest, are first divided into parts (e.g. Chapter 2, part I. Animal communication) and only then into sections (Chapter 2, part I, 1. General debates. Signals, not words) and subsections. By this token, whenever content from a distant chapter is brought into discussion, its exact location will be indicated.

CHAPTER 1. LAYING OUT THE PROBLEM [OR, THE CHICKEN AND THE EGG, SHORT VERSION]

In this first chapter, I will present the problem that motivated this thesis, and provide rationale for its undertaking. In order to do so, it will be necessary to introduce three notions: music, musicality, and musical enculturation. Unlike chapters 2, 3 and 4 where I will scrutinize matters in further depth, in this chapter the level of complexity of the concepts to be introduced will be limited by the goal of the chapter itself: to triangulate a first understanding of the problem. By the end of this chapter thus, we should be able to grasp the problem in a first level of complexity, as well as sketching preliminary research questions.

1. [The chicken] What is music?

Music seems to be a human universal that exists in every known culture and emerges spontaneously in an early stage of human ontogenetic development (Vitouch & Ladinig, 2009). Although claims concerning its universality have been often based on assumptions and extrapolations (Ilari, 2016), recent large-scale systematic research has found that indeed some form of music appears in every assessed society (Mehr et al., 2019). In any case, the simplicity of discriminating music from 'not-music' relies heavily on one condition: that the person and the music in question are products of roughly the same society. In order to have a better sense of how our notion of music is shaped by our culture, and how such a notion might clash with those generated by other cultures, let us first examine the most familiar example— music within Western societies —and then problematize the matter by adding the alternatives and nuances that other, non-Western cultures provide.

In the case of contemporary Western societies that have an increasingly global reach, music is conventionally conceived as *"…a consumable commodity constituted of complexly patterned sound that is produced by a class of specialists and engaged with through listening* *for primarily hedonic reasons"* (Cross, 2007, p.1). To put Cross' definition in other words, for a vast number of people in the planet, music is something to listen to, a primarily aural phenomenon. Western music is mostly heard in pre-recorded forms through devices that allow its playback at will, after being previously composed by someone, in most cases following the rules of tonality. It can be the absolute centre of attention as in the case of a concert, or mere background as in an elevator, restaurant or mall. Music in Western societies is engaged with mostly for its pleasurable effects— dancing, relaxing, sharing, etc.

All of these characteristics sound most probably familiar and perhaps even evident to the reader, since the notion of music presented above has permeated a large proportion of the world's population. However, things get more complicated when taking into consideration what music is— or is not —for people in non-Western societies. Indeed, as ethnomusicology has shown, music varies from society to society— both in terms its structural features and the functions that it may fulfil —to the extent that the music of a given culture may not be recognizable as such by members of a different one (Ibid.). It was previously stated that music is a human universal that exists in every known culture. Although the latter is definitively true from a certain point of view, it is more precise to say that 'music' can be found in every known society given that trained musicologists have had the breadth of knowledge and the emic sensitivity to recognize it as such, and not because the phenomena studied necessarily presented any evident similarities with Western standards (Blacking, 1995). In fact, some societies can be found that have no actual word for music, or whose concept of it is accorded a significance quite unlike that more widely associated with the word 'music' (Blacking, 1995).

Another level of complexity to be added is that the clear-cut Western distinction between music and other forms of communication such as language may not be so elsewhere. A classic example is that of the Amazonian Suyá, who have built four related yet distinct-enough categories (Seeger, 2004). The Suyá have *ngére* (roughly equitable to 'song') and *kapérni* (roughly equitable to 'speech'), but also *sarén* (telling) and *sangére* (reciting), which are not subcategories of any of to the former two. Ngére presents a priority of melody over text— time as traditionally found in Western singing —, with both text and melody being set by a non-human source. In the case of *Kapérni*, text has priority over melodic aspects, with both text and melody determined by the speaker themselves. *Sarén* and *sangére*, the more relevant categories for our current purposes, exhibit a relative priority of relatively fixed texts over relatively fixed melodies. Accordingly, they cannot simply be put under the umbrellas of 'speech' or 'singing'. Because human communication capacities are subdivided into different numbers of communicative media depending on the culture (i.e. Western dichotomic language/music subdivision v/s Suyá triadic subdivision), the sum of such capacities has been conceived as a universal communicative 'toolkit' (Cross & Woodruff, 2009).

Yet another manner in which the 'musics' of the world differ is in terms of participation— who is allowed/expected to make music, and who is not. In this respect, American ethnomusicologist Thomas Turino classified music into two modes of performance: participatory or presentational (Turino, 2008). In participatory music, all attendants are allowed or expected to actively do something, be it play an instrument, sing or and/or dance. Presentational music, on the other hand, is characterized by a clear divide between— a few, expert —performers, and the audience. Audience members might dance or sing along but they remain out of the focus of attention. While to date in many— perhaps most —non-Western cultures music mainly involves overt action and active group engagement (Arom, 1991), Western societies have progressively privileged the presentational mode (Cross, 2013). Thus, while music jams and campfire singing do occur, classical, rock or jazz concerts are the most commonly appreciated forms of music, and the ones recorded and listened to by the masses through smartphones and other devices. As in the case of the tendency to privilege solitary listening, the communicative character of music and its social character, although compelling and widespread in contemporary literature, remain to some extent assumptions (Ilari, 2016).

Many more elements that set Western and non-Western forms of music apart— or together —could be introduced and explained in this section. Furthermore, rhythmic and/or melodic behaviour performed by other species such as birds (Emery, Seed, Von Bayern, & Clayton, 2007; Tierney, Russo, & Patel, 2008; Wilson & Cook, 2016) and cetaceans (Ford, 1991; Rendell & Whitehead, 2001), as well as a wealth of musicological discussion could contribute to the widening of what 'music' might stand for. However, those so far presented should provide the reader with the necessary context and nuances when the moment comes to lay out the main thesis. Fortunately for the reader, both for theoretical reasons and due to the population that participated in the three studies that will be presented, in this work we will focus on the largely intuitive, Western definition of music that has been already outlined. Nevertheless, such a focus will not take Western music's characteristics for granted, as we will keep in mind how things could perfectly be elsewhere. Because I chose to focus on a Western conception of music (and therefore of musical enculturation), a myriad of other alternatives scattered around the world are thus inevitably left aside. For this reason, it can be said that our quest privileges psychological over anthropological or ethnomusicological interests, or that I will privilege depth over universality.

More concretely, the contents in this section will provide nuance and depth when assessing how children's engagement with music changes over time, becoming progressively more 'Western'. In particular, the ways in which— contrary to most known non-Western cultures in the planet, and to all of human culture until the introduction of recording technologies in the nineteenth century —children begin to autonomously listen to pre-recorded, exclusively aural music.

2. [The egg] What is musicality?

As discussed in the previous section, the body of elements involved in the question of what defines music has displayed a continuous growth. Given that a definition of music can and has been (by some) stretched to include all studied human societies, types of sound, noises, and even plain silence (Davies, 1997), an alternative approach to the question has been to enquire what are the capabilities that allow us to perceive and distinguish music from other kinds of stimuli in the first place, whatever our culture. Such a capacity has been referred to as 'musicality' (Honing & Ploeger, 2012; Malloch & Trevarthen, 2009; Trehub, 2003, 2010).

In this section, two main related yet distinct approaches will be heuristically characterized: musicality as the seed of music, and musicality as the seed of human communication (music included). Equally importantly, three main ideas will be highlighted. First, that musicality is, at least partially, an innate capacity. Second, from a developmental point of view, there is a causal relationship between musicality and music: the former enables the latter. Finally, infant-directed prosody bears musical qualities intimately related to early musicality, and therefore to eventual engagement with the cultural product we refer to as music.

2.1. Musicality as the seed of music

Musicality has been used in the literature with different levels of conceptual precision, and the same can be said about its relationship to music. A general definition of musicality has been proposed, as a "sensitivity for music, predispositions for processing musical experiences, and expression of musical skills" (Papousek, 1996). A next level of precision is that provided by Honing and Ploeger, who define musicality and its relationship to music as *"a natural, spontaneously developing trait based on and constrained by our cognitive system,* *and music as a social and cultural construct based on that very musicality. Without musicality no music."* (Honing & Ploeger, 2012). Such a definition highlights the cognitive nature of musicality (see the 'music as cognition' hypothesis, Honing, 2011; Serafine, 1983), consequently enquiring about the precise cognitive functions that can be shown to be fundamental in developing musical behaviour. Although such functions have not been completely elucidated or agreed on (Trainor, 2006), relevant candidates posited in the literature are relative pitch, tonal encoding of pitch, beat induction, and metrical encoding of rhythm (Honing & Ploeger, 2012).

It is not necessary for us to define and delve into each of such cognitive abilities. However, beat induction is a case the characteristics of which will help us building a sense of musicality as a whole. Beat induction is the cognitive skill that allows us to hear a regular pulse (i.e. the 'beat' or 'tactus') in music, to which we can then entrain (Patel, 2008). Perceiving such a regularity in music subsequently enables us to perform a number of musical behaviours, from the simplest ones likes nodding or tapping along to the more sophisticated, like dancing to the music or playing in a jazz ensemble or classical orchestra. Remarkably, it is not the complexity but rather the simplicity of beat induction that makes it the most relevant, for it is perhaps that simplicity that allows it to emerge sooner than other, more complex elements of musicality such as, for example, tonal encoding. Some researchers first considered beat perception to be acquired throughout the course of the first year of life (Hannon & Trehub, 2005), and suggested that the infant's development of the sense of a pulse would be foremost developed through being rocked to music by their parents. The precocity of beat induction has since been proved to be even greater, as recent empirical studies show that beat induction is already functional in newborns and young infants (Winkler, Ha´den, Ladinig, Sziller, & Honing, 2009; Zentner & Eerola, 2010).

The precocity of beat induction— from the very outset of life —has been taken as evidence of its innateness, rather than the result of learning (Winkler et al., 2009; Henkjan Honing, Ladinig, Háden, & Winkler, 2009). Given that beat induction is but one component of the set of cognitive abilities that musicality comprises, its innate character illustrates in turn how early in life musicality can be considered to manifest itself.

2.2. Musicality as the seed of human communication (music included)

The precocity of musicality is not only supported by the literature that considers it a cognitive trait that enables our engagement with music. A related yet different field of research considers musicality to be at the heart not only of musical behaviour and cognition, but also of early human communication. By implication, musicality is thought by some to be the root of the whole of our communicative capacities, language included. Perhaps the most widespread conception of musicality in this sense, is that presented by Malloch and Trevarthen:

> "[musicality is] …the expression of our human desire for cultural learning, our innate skill for moving, remembering and planning in sympathy with others that makes our appreciation and production of an endless variety of dramatic temporal narratives possible – whether those narratives consist of specific cultural forms of music, dance, poetry or ceremony; whether they are the universal narratives of a mother and her baby quietly conversing with one another; whether it is the wordless emotional and motivational narrative that sits beneath a conversation between two or more adults or between a teacher and a class. (Malloch & Trevarthen, 2008).

This definition is evidently more complex than the ones above presented, not so much because of it being comprehensive per se, but because of the several dimensions it aims to bring together. Unlike Papousek, or Honing and Ploeger's definitions in which musicality orbits predominantly around one single notion— music —, Malloch and Trevarthen's conception adds communication as an inherent dimension.

Because of their scope, the authors went beyond musicality itself and coined the concept of ‘communicative musicality’, which refers to *the ‘co-operative and co-dependent communicative interactions between mother and infant’ (Malloch, 2000)*. In this view, the building blocks of musicality correspond not so much an individual’s cognitive traits, but to behaviours displayed in the context of and in order to enable interpersonal dynamics. Communication between mother and infant is a dialogue shaped by three main musical elements: pulse, quality and narrative. Narrative is a more complex construct founded in pulse and quality, that allows interactants to share a sense of passing time, and to create and share emotional experiences. Quality roughly corresponds to the contours of expressive vocal and body gesture, which are more commonly known as the infant-directed register. The infant-directed register and its main vocal manifestation, infant-directed speech, will prove to be of first importance in this work as they convey the idea that there is a musical, melodic nature to the infant’s and— especially —the mother’s prosody.

Malloch and Trevarthen’s approach has received at number of criticisms. A first one concerns the ambition and complexity of the approach. As previously discussed, conventional understandings of musicality involve a sensitivity for, understanding of or talent for music (e.g. Papousek, 1996). It could therefore be expected for communicative musicality to consist of the sharing of such sensitivity, understanding or talent with others. The fact that, on the contrary, so many more elements and layer of complexity are added to the definition of musicality and communicative musicality make the latter confusing and even ‘rambling’ (Trehub, 2010). A second criticism concerns Malloch and Trevarthen’s assumptions. Their proposal rests on the supposition that infants and mothers possess innate intuitive access to each other’s subjective states, such access in turn enabling reciprocal emotional and behavioural coordination. These ideas, partially built on psychoanalytic foundations, deviate— both in theory and standards of evidence —from mainstream approaches to the psychology of music (Ibid.). Finally, their work has been interpreted as establishing and reinforcing an idealized image of mother–infant interaction (Young, 2005). Their theory would subtly present a normalizing standard to which all mother–infant interactions should aspire, leaving little room for diversity of parenting style, let alone considering versions of parenting practice among different communities or cultures.

Still, Malloch and Trevarthen's notions of musicality and— particularly —communicative musicality have been widely welcomed in academia and there is a significant number of works that take into account such a notion (Barrett, 2011; Dissanayake, 2008; Ilari, 2005; Van Puyvelde et al., 2010). Even if their assumptions, background theory or methods may be problematic, their main idea— that musicality lies beneath not only human engagement with music, but human communication in general —is largely compatible with other major approaches to early musical behaviour and communication (Dissanayake, 2000, 2001; Falk, 2004; Falk, Rathcke, & Dalla Bella, 2014). As an eloquent illustration, even though Trehub (2010) criticised the inherent association between musicality and communication, she later on stated upfront that early caregiving *is* musical (Trehub, 2016).

As stated in at the beginning of this section, before moving on we should finish bearing three ideas in mind. First, that musicality has and is conceived in different manners which, nevertheless, share some common elements: musicality is a human capacity some elements of which can be observed practically from birth, particularly those involving beat or pulse. Second, from a developmental or ontogenetic point of view, there is a causal relationship between musicality and all consecutive engagement music: the former enables the latter— as Honing and Ploeger (2012) put it: without musicality no music. Last but not least, infant-directed prosody bears musical qualities intimately related to early musicality, and therefore to eventual engagement with music —the latter being, as discussed in section 1, a culture-specific product. The literature supporting this point is well documented and theorized, to the point that it will require over two chapters to understand it in light of the full available evidence. Because infant-directed speech is at the same time a form of biological communication and an attachment device, it will prove crucial in explaining not only the cognitive capacities necessary for engaging with music, but also the *motivation* for doing so.

3. [The hatching] Musical enculturation

But let us not get ahead of ourselves. After dealing with music and musicality, there is one last notion to be introduced: musical enculturation. This notion can be said to be more important than the previous two insofar it concerns their interrelation, being precisely this interrelation that will be scrutinized and around which all remaining chapters will revolve one way or another.

I have so far depicted music as a cultural product, a wealth of musical practices the specificities of which are defined by each society. I have also characterized musicality, the sum of capacities that enable human beings to perceive and engage with music. By this token, on the one hand, we have an infant born with all the potentialities that musicality grants and on the other, adults that have built a notion of, and forms of engagement with music proper to their society. The question then remains, what happens in between? Music and musicality are interrelated in an ancient cycle in which they complementarily enable each other: music exists in a given society, new members are born into that society and— by means of their musicality —absorb the available forms of music and musical practices until these are naturally integrated, to eventually transmit them to a third generation and so on and so forth. Therefore, the process by which a human being uses musicality to naturally and informally absorb its culture's music represents a form of enculturation.

Originally conceived by anthropologists, the term 'enculturation' concerns much more than music alone and extends to literally all aspects that characterize a given culture. It can be defined as:

> "…the process where the culture that is currently established teaches an individual the accepted norms and values of the culture or society where the individual lives. The individual can become an accepted member and fulfil the needed functions and roles of the group. Most importantly the individual knows and establishes a context of boundaries and accepted behaviour that dictates what is acceptable and not acceptable within the framework of that society." (Kottak, 2017)

This definition may well sound either too general or too focused on morals for a thesis centred on music. However, it reflects the fact that absorbing our culture's music does not happen in the void. As such, Kottak's general definition of enculturation, besides providing us with an encompassing notion of the term, will be important when first laying out the problem in the next section, and again when doing so in full complexity in chapter 5.

In any case, for the purposes of this section, I shall rather focus in the literature of music psychology, where enculturation is often used without formal definition and mainly in contrast to formal training (Corrigall & Trainor, 2014; Grannan-Rubenstein, Grannan-Rubenstein, & Thibodeau, 2014; Kreutz & Feldhaus, 2018). An explicit definition of musical enculturation was provided by Hannon and Trainor (2007), who defined it as *"…the process by which individuals acquire culture-specific knowledge about the structure of the music they are exposed to through everyday experiences, such as listening to the radio, singing and dancing."* (p.466). Thus, while enculturation consists in basic musical capacities being 'passively' modified by everyday experience within a particular culture, in formal musical experience perception and production skills are 'actively' trained to a high level, musical knowledge thus becoming explicit (Hannon & Trainor, 2007). For instance, most people in Western societies roughly know how to tell when someone sings out of tune, even though they have not been formally trained to do so. Accordingly, it can be said that such a capacity was learned by means of slow, mainly subconscious, comparatively effortless enculturation.

4. Formulating the problem

The goal of the previous sections in this chapter was to provide the minimum conceptual substructure that will now allow us to build the rationale for this dissertation. I will do so by proceeding in three steps. First, I will present some of the defects and loose ends that the notions of music, musicality and enculturation exhibit to date. I will highlight lacunae in the contemporary literature that have prevented some research questions from being formulated as well as obstructing the formation of connections between different theoretical and empirical findings. Second, I will explicitly formulate some research questions that need to be taken into consideration in order to bridge these theoretical or empirical gaps, and deal with the loose ends. Third, I will sketch my suggested approach to answering these questions. Finally, I will briefly state what could be gained by answering these novel research questions.

Let us thus begin by providing a succinct critique of the literature. I have established that music and musicality are socio-culturally interrelated by complementarily enabling each other. I have also established that— from an ontogenetic point of view —there is a sequential relationship between the two: musicality is either innately endowed or early acquired by the child, and it only eventually translates into musical behaviour as established by its to-be culture through a process called musical enculturation. Children thus progressively narrow-down and selectively develop their music perception and behaviour, so as to match their surrounding cultural expectations.

However, it is my sense that theorization and research on musical enculturation in Western societies have heavily focused on cognition— notably, the perception of pitch and rhythm. As a result, the social and socio-cognitive mechanisms that support the emergence of such cognitive skills in the first place, are less considered in the field of early musicality (Ilari, 2016). Furthermore, other relevant aspects of music that are overtly or tacitly imposed on children have been left largely unattended, or perhaps not even been conceived of as a form of enculturation. One such aspect is the widespread practice of listening to pre-recorded music, and the sources of motivation that Western society provides to do so. As previously discussed (section 1), listening to pre-recorded music is one of the paramount and distinctive features of Western music. It strongly implies a presentational mode of music engagement, in which the performer (having previously recorded the music) cannot possibly take into account the audience's potential reactions, and the audience's (the listener's) only chance for activity resides in inner, mental forms of participation (Cross, 2010). Aside from the level of activity, listening to pre-recorded music often additionally implies doing so alone (i.e. without interacting with anybody in doing so), and often with no multimodality involved but attending exclusively to sound. As previously indicated, although all of these aspects of Western music listening may seem obvious and irrelevant, they no doubt depend on musical enculturation and need to be learned somehow, at some point.

In other words— and this will be one of the pivots of this dissertation—, the assumption that an infant or child will innately pay attention to recorded music the way an adult would (in a sustained, autonomous manner, with no mediating interaction and unimodally by means of hearing alone) is, as will now be argued, a misleading one.

Indeed, the practice of autonomously listening to pre-recorded music is crucially distinct not only from other non-Western cultures around the globe, but from the most relevant early musical experiences: infant-directed speech and infant-directed singing— both multimodal and inherently interactional. Such fundamental differences can be conceived of as crucially problematic in the light of the fact that a continuity between the infant-directed register and all subsequent engagement with Western music has to date been largely taken for granted yet not accordingly explained. When placing the origin of all future music-related behaviour in early interaction (Malloch & Trevarthen, 2008), or considering early caregiving as musical (Trehub, 2016) and infant-directed speech as 'quasi-music' (Trehub & Degé, 2016), a causal relation is drawn more or less tacitly: that one is necessary for the other to take place. Yet, as will be discussed (chapter 5, section 4) while early interaction, and interaction within the infant-directed register take place during the first years of the child's life with a peak around 6 months of age, the most common form of Western music engagement (listening to a music recording) requires further enculturation and is not obviously observable until the end of childhood/early adolescence. By this token, a mechanistic, observable chain of events linking the two phenomena is tacitly assumed while remaining largely unquestioned, unproblematized, undescribed, and unexplained.

Let us provide a further illustration of this omission. The two most relevant early musical experiences— infant-directed speech and infant-directed singing —are inherently interactional and take place in the context of a (more often than not, significant) relationship. They are both also multimodal: the singing or uttering voice is not heard alone, but organically orchestrated with the utterer's facial expressions, gestures, touch and even their smell. As such, infant-directed speech and infant-directed singing are naturally compelling. They are both performed live by parents, often face-to-face to the child and also often from a short distance. Therefore, infants and children pay attention to them for as long as they take place, as much as any of us would in the same circumstances. After all, who could ignore a private concert performed especially for us? Who could ignore speech or song vocalized from a few centimetres' distance? Accordingly, as we shall see, an important part of the literature on early musicality and musical behaviour focuses on forms of musical interaction (e.g. joint and supported singing, dancing, jointly doing actions, etc.) and the concomitant interpersonal synchrony that emerges in them (see chapter 3 section 1.4 and chapter 5 section 3.5).

The interpersonal, interactional, immersive quality of infant-directed speech and infant-directed singing naturally and strongly motivates its audience's attention, be it young or old. A very different case is that of recorded music which, as a cultural product, presents further challenges to an infant or a child in terms of how to approach it or what to expect from it. In other words, recorded music is opaquer to a child than IDSi or other forms of interaction, opacity being a term that I will discuss in further depth (see chapter 3, section 1.3 and chapter 4, section 2). The opacity of recorded music renders the latter's appeal and value not as straightforward, and evidently more meagre when put next to the charm of the comparatively-more instinctive and immediately-rewarding infant-directed register. Although it is true that almost from birth infants have the capacity to recognize and entrain to the pulse of recorded music (see Chapter 5, section 4.1) they only do so for around 10 seconds, quickly losing interest. Effectively, a unimodal (aural), presentational (not interactional) recording of a singing voice does not carry the ultimate appeal of human interaction, this being even more the case of purely instrumental music.

At this point some questions stem naturally from the argument. Some very relevant ones have already been put forward by Beatriz Ilari (2016) in an analysis of the social components of early musicality. Here are some examples:

> "while musical engagement in the early years has often been described as a form of its own (Bjorkvold, 1992; Trevarthen et al., 2014), in what ways (if any) is it related to *culture-specific* forms of collective music making, like singing or playing in groups later in life? (p.34, emphasis added)

> … a question that needs to be asked is when and how music begins to be less about "itself" (Malloch & Trevarthen, 2009) and musicking less about child culture (Bjorkvold, 1992), becoming more constrained and in accordance with the tenets of adult culture and aesthetics. For example, when and

> how in the course of development does the distinction between presentational and participatory forms of music making (Turino, 2008), that is blurred in early childhood (Young & Ilari, 2012), become more evident? (p.33)

> Aside from formal schooling, what other forces compel humans to move away from playful and collaborative forms of musicking (e.g., musical play) to more *rule-governed* forms that *conform to social and cultural norms?* In other words, when, how, and why do our *conceptions of music become more constrained* in the course of human development? (p.34, emphasis added)

The reflections contained in these quotes overlap with the arguments offered in the present section. Ilari herself contemplates the social aspects of early musicality through the prism of enculturation, therefore using words like 'culture-specific', 'constrained', and 'norms'. Nevertheless, her scope remains different from the one that structures this dissertation. For instance, while in the first quote her interest is directed towards evidently-social activities such as collective music making and singing or playing in groups, in the present work I will focus on another culture-specific form of engaging with music: autonomous listening to recorded music. Similarly, I will detach from Ilari's contemplation of general social musical interaction by instead circumscribing possibilities to those that hold the IDReg at their very centre.

Accordingly, let us now suggest some new questions according to the contents in this chapter: where does a child's motivation for listening to recorded, non-infant-directed Western music come from? After having experienced the luxury of infant-directed speech and infant-directed singing, why would a child settle for a paler alternative such as a recorded music track? At which moment(s) of development does proper attention to recorded music starts taking place (beyond a few seconds)? Can some kind of continuity be observed between children's initial motivation to engage with infant-directed speech or infant-directed singing, and their later motivation for engaging with recorded Western music? Are there any extra-musical phenomena, that could help explaining such a continuity? In consequence, can some kind of mechanistic, observable chain of events link early musicality and later listening to recorded Western music?

As will be discussed when properly presenting the main thesis, efforts have already been made in the direction of these questions. However, such efforts focus on interaction and synchronization in a broader, more general scope than that concerning the use of IDReg. As an alternative, the present dissertation will focus on the IDReg as an element of continuity. Such a focus is— partially —arbitrary. It is not arbitrary insofar the IDReg has already been linked to later engagement with recorded music, and insofar a vast amount of literature supports the register's place in human communication and attachment (see chapter 4, section 3). What remains arbitrary is that I choose to focus on it knowing that there are further, no less important concomitant elements such as further forms of interaction and interpersonal synchrony. In this sense, I will willingly and purposively choose not to focus on other forms or aspects of interaction that can be found in existing literature. As a consequence, what the present dissertation will return, at best, is a better-informed sense of the relative importance of the use of the IDReg as one of many elements that have an impact in musical enculturation.

The mentioned chosen focus on autonomous listening to recorded music also implies a rather analytical approach. Indeed, the main thesis will aim to distinguish and isolate the impact of the IDReg from that of other, concomitant factors, and attention to recorded music from other musical activities that have been comparatively-better documented. Introducing such distinctions holds the promise of being able to distil the influence of IDSp and—hopefully —introduce novel nuances to our understanding of early musical activity and the motivation that lies beneath it. On the other hand, dissecting phenomena into units such as recorded music and aiming to isolate them from other phenomena such as interaction risks becoming a rather artificial exercise. Such a choice of an analytical approach will be re-discussed in light of this dissertation's findings in the final discussion.

It seems that much would be gained from answering the novel questions hereby presented. For a start, it would bring insight into preceding, tangentially related questions (such as, for instance, the ones above quoted). At a deeper level, however, the goal of science is to generate theory by testing it empirically and systematically. As research in a field progresses, evidence can and must not only be accumulated, but integrated— if two or more related phenomena are coherently connected, they validate each other and well as making the field stronger as a whole. In this particular case, the research field is the psychology of music, where there has been research focusing on the musicality of the infant-directed register on the one hand, and in children's engagement with Western music on the other, but the relationship between these two research topics has been only implied and not explicitly scrutinised. As compelling and apparently plausible as such a relationship may be— and I think it is —, not only has it not been explained but no attempt has been made to explain it, thus running the risk of accepting an unexamined and unacknowledged assumption instead of generating theory at higher explanatory levels.

This thesis is thus an attempt to contribute to the field of music psychology by explicitly addressing the presented omissions. As announced in the introduction, the start point for such an enterprise will be to understand in as much details as possible the main element in early musicality: the infant-directed register in general, and infant-directed speech in particular. In turn, in order to truly understand the scope, features and functions of the infant-directed register (chapter 4), two previous notions will also have to be reviewed in depth and critically addressed: human communication (chapter 2) and attachment (chapter 3). Once we achieve a state-of the-art understanding of the infant-directed register I will (in chapter 5) re-appraise the problem in its full intricacy, generate our thesis, and lay out a research programme.

CHAPTER 2. COMMUNICATIVE USE OF SOUND IN ANIMALS AND HUMANS

The main thesis in this dissertation addresses dynamics involving IDSp, and their potential impact on toddlers' motivation to listen to recorded music. Before plunging into the linking element of the thesis (IDSp) and its motivational aspect— attachment—, I will first address a common status of speech and music: communicative sound.

There are of course many angles from which to approach the perception of speech and other communicative sounds, however, when considering that the target recipient of both kinds of sound corresponds to small children still early on their way to acquiring language, choices narrow. As we shall see in the second part of this chapter, the use we humans make of sound at the verbal and non-verbal levels is different. Spoken, uttered language comprises several layers, at times radically different in nature. One consists of a complex referential and combinatorial system practically exclusive to human beings in its collective use. The other, contained in non-verbal aspects of communication such as prosody, makes a use of sound that, far from exclusive, is common to a vast number of non-human animals.

The target population of this dissertation are toddlers that cannot yet properly speak and have so far communicated and socialized mainly through non-verbal means. Because of this, imposing a logocentric point of view to the study of their relationship to surrounding speech— and music —would be misleading. In other words, studying the manner in which toddlers open to cultural products such as speech and music must take into account the manner in which they have so far dealt with social sounds in general, before distinguishing them as one *or* the other. This first part of the present chapter will thus aim to provide a minimal, necessary understanding of communication, beyond its human forms. Counting on such an understanding will subsequently allow us to tackle human communication including its other relevant layer: language and its referential/combinatorial quality. By the end of the chapter, we should be able to understand from a communicative point of view *how* children deal with sound before entering language, and the revolution the latter introduces in their social lives. This will in turn put us in a position to consider the role of communication in attachment and have a better sense of *why* children would be motivated to perceive or generate sound in social contexts in the first place.

Part I. Animal communication

Unless one can postulate a radical discontinuity between human and non-human communicative capacities— difficult in the light of evident evolutionary continuities —the probability that human modes of communicative interaction relate in a principled way to non-human modes, particularly those of non-human primates (as our closest relatives) must be acknowledged.

Animal communication is a matter traditionally and mostly dealt with by ethologists and comparative psychologists. Relying on the theory of evolution, these disciplines focus on the role of communication in survival and reproductive fitness. In other words, forms of animal communication have been largely described regarding their role in attracting mates, outsmarting predators or prey, and obtaining attention from caregivers, to mention a few relevant dynamics. Although these dynamics may at first seem rather far away from the human infant-directed register, they will in fact help us understand the latter's status as a biological signal, thus demonstrating its biological relevance.

When approaching animal communication, one of the main goal of scholars has been to generate an understanding of communication as a general category applicable to all communicating organisms (Harms, 2004). One such a basic theoretical tool is the idea of a 'signal'. Animal communication is widely understood in term of signals, and how these signals may contribute to an animal's welfare, be it being sociably successful or simply surviving from predators.

In this first part, three main general aspects of animal communication will be introduced: whether it will be more useful understanding animal signals as carrying information or influencing behaviour, the evolutionary character of animal signals, and the 'honesty' of signals, as problematized by the 'handicap principle' and the concept of 'Machiavellian deceit'.

1. Signals, not words

1.1. Animal signals as influence

In traditional ethological theory, signals have been considered basically as stimuli that trigger behaviour in other organisms. For a start, a well-known definition of animal signals is the following:

> 'a behavior or phenotype produced by one individual (the signaller) that serves to influence the behavior of a second individual (the signal receiver) by transmitting information.' (Lachmann, Szamado, & Bergstrom, 2001, p. 3).

A definition of this kind has proven fit for describing and understanding communicative situations. However, in its use among animals other than humans, a consensus has not been reached as to how exactly do signals trigger behaviour in their recipients. It is not clear either whether such signals generate any sort of 'meaning' (Cheney & Seyfarth, 2010; Oller, 2004; Rendall, Owren, & Ryan, 2009) or— more important for our current purposes —whether they convey information.

Most authors that approach animal communication (or language) from an informational point of view, avoid an explicit definition of the term 'information' or else are ambiguous in their use of it (Ibid). Indeed, authors tacitly use the construct following the understanding provided by Shannon & Weaver's (Shannon & Weaver, 1949) Information

Theory, in which ‘information’ constitutes a change (increase or decrease) in the level of uncertainty in respect of an aspect of the environment. However, they often fail to make this clear, conflating what is a technical, mathematical definition with the vernacular definition of "information".

Although understanding signals as carrying information has proved useful to some, it also raises a number of issues. For instance, although animal listeners often respond to vocalizations ‘as if’ they contained information— a phenomenon referred to as ‘functional reference’ —, callers prove to be fundamentally unaware of the informational value of their own signals as well as largely demonstrating a surprising absence of the intention to ‘inform’ (Rendall et al., 2009). Another issue is that information-based approaches to animal communication often go one step further by relying on the way information has so far been treated in language theory, assuming that signals encode information that ‘refers’ to states of affairs in the environment. Nonetheless, while some research has shown quite precisely the extent to which primates do behave as if their vocalizations are referential (Cheney & Seyfarth, 2010), Owren and Rendall (2001) warn that such a ‘metaphor-as-explanation’ approach has distanced many researchers from normative biological perspectives and the structure-function relationships that are likely to exist in animal signals. Owren and Rendall also deem it problematic to adopt an *a priori* language-based view when approaching nonhuman primate vocal behaviour.

Characterizing signals as having encoded, referential information raises a third issue, as it implies privileging a metaphorical approach over mechanistic or functional ones. Yet accepting that information may play almost no role in animal communication implies that the semantic information conveyed by human words has no parallel in the vocalizations of any nonhuman creature, a conclusion in all likelihood premature at best (Seyfarth & Cheney, 2003). Seyfarth and Cheney illustrate this problem pointing that any information-based definition of language (where meaning is still, somehow, conveyed) has ultimately failed in its logical-propositional endeavour:

> 'In language, where the use of sounds to represent features of the environment is no longer in doubt, the question "What do words mean?" is both fundamental and unresolved. As Wittgenstein (1953), Quine (1960), Putnam (1975), and others have argued, it may never be possible to state precisely what an individual means— and what information a listener acquires— when a speaker uses a particular word.' (Seyfarth & Cheney, 2003 p. 164).

In other words, given that theories of language themselves find it hard to explain what the relationship between meaning and information is, it is premature to assume that there is no informational content in animal communication.

Largely in an attempt to comply with basic scientific and evolutionary principles, researchers have looked for alternative approaches to communication in general and animal communication in particular when addressing the issues surrounding information. The main alternative has been the notion that the function of signalling is to *influence* the behaviour of perceivers rather than transmitting any language-like information (Guilford & Dawkins, 1991; Owings & Zeifman, 2004). Such an approach highlights links between acoustic structure and vocal function, particularly if senders are assumed to use whatever calling strategies allow them to influence receivers (Owren & Rendall, 2001). Corollaries include emphasizing the role of signal structure in exerting influence and expanding the conception of communication beyond representational exchanges.

Influence-based theories highlight the role of signals' structure, opening in turn a vast field of exploration on the morphophoric, or form-bearing properties (Attneave & Olson, 1971) of acoustic signals. Motivational-structural rules of animal communication (Morton, 1977) can perfectly be included among such morphophoric properties. The concept, first pointed out by Collias (1960), consists of the structural convergence of many animal sounds (mainly birds and mammals) used in 'hostile and 'friendly' contexts, which suggests a relationship between sound structure and function:

> 'Simply stated, birds and mammals use harsh, relatively low-frequency sounds when hostile and higher-frequency, purer tone-like sounds when frightened, appeasing, or approaching in a friendly manner. Thus, there appears to be a general relationship between the physical structures of sounds and the motivation underlying their use.' (Morton, 1977, p.855).

Such rules favour the use of communication instead of (or in conjunction with) fighting to attain resources, both at intra-specific (within a species' repertoire) and inter-specific (between different species) levels.

Influence, structure and motivation are also closely related to assessment-management theory (Owings & Morton, 1997). Assessment involves generating self-interested adjustments to circumstances based on the extraction of signal-based clues from other individuals (and their contexts). Management involves equally self-interested efforts to maintain or change current circumstances by regulating the behaviour of others (regulation is close to management and related to influence) by means of signals. Morton's theory relies on indexical information, which tells us something about the source of the signal, not just the signal itself. All of these influence-related concepts will facilitate our understanding of parents and children's vocalizations later on in this dissertation.

Thus, although information is an extremely useful construct when coding or writing, as discussed in this section, its application remains controversial and often inconclusive when approaching sentient beings whose cognitive and social dynamics are not comprehensibly understood. We may well know exactly what information is being sent via telegraph or the through the internet given that the governing rules of such communication have been explicitly defined. However, such certitude is not present when studying other animal species— or pre-verbal humans infants. On the other hand, it is possible to influence through communication, influence being an observable. Because some of the later, empirical chapters in this dissertation will precisely assess the behaviour of children unable to express themselves verbally, an influence-based approach to non-verbal communication will prove more useful. I will use 'information' (in quotes) when referring to communicative elements that reduce the receiver's degree of uncertainty, although we cannot say what such information consists of exactly.

1.2. Signaller and receiver. Animal signals as evolved influence

Intimately related to the information versus influence debate, the added complexity of contemplating not one but both 'ends' in a linear model of communication— the signaller and the receiver —has also proven to be highly relevant and will allow us to refine our definition of what an animal signal is.

Given that a significant number of cases for animal communication prove to be understood by considering the signaller and the signal's features only, responses to vocalizations have been often thought to be fixed and invariant (Wallman, 1992). However, a growing number of cases prove to be less simple, especially in nonhuman primates [NHP]. For instance, an absence of the intention to inform by calling animals has already been mentioned; although listeners sometimes respond to vocalizations 'as if' they contain information, callers have proven to be fundamentally unaware of the informational value of their own signals (Rendall et al., 2009). In the same manner, even if animal alarm calls comprise acoustic features that may grant them an attention-getting and arousing quality, acoustic features alone ultimately fail to make sense of several communicative instances (Seyfarth & Cheney, 2003). For instance, even if the morphophoric features of a given signal (from a motivation-structural point of view) seem clear, the possibility of the recipient failing to acknowledge what the signal reveals of the producer cannot be excluded. Neither can the possibility of more sophisticated inference coming into play (Owren & Rendall, 2001). It is thus crucial to separate a given vocalization and its intended purpose, from potential subsequent cognitive processing performed by listeners that may or may not allow the latter to draw some sort of inference.

It is no less crucial to separate a given vocalization and its intended purpose, from the signal actually generating the influence it sought to generate in the recipient, the purpose of a signal being determined by its evolutionary history. Thus, to be more precise, in the present dissertation we will understand an animal signal not just as influence, but as one determined by evolution (Smith & Harper, 2003):

> 'any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved' (p.3).

By adding such an evolutionary layer to the definition, any act or structure that has an incidental effect may be considered a cue, but not a proper animal signal. For instance, a leopard's sharp teeth may act as a cue of ferocity to a smaller animal hiding in the bushes. The influence that such sharp teeth had on the smaller animal was that the latter made sure to escape the leopard's senses, which is certainly not the effect they evolved for. On the other hand, the same sharp teeth may deter competing males, which is something they did evolved for. In such a case, they are indeed acting as a full-fledged, animal signal. Accordingly, whereas animal signalling may occasionally fail, we will understand biological communication exclusively as the successful *completion* of a signalling act (Scott-Phillips, 2008), where an evolved signal triggers the response it evolved for.

The role of the receiver is thus a necessary element in Smith and Harper's conception of animal signals and Scott-Phillips' conception of biological communication, where an evolved feedback system is in place. The idea that animal signals work through evolved dynamics between signaller and receiver will be crucial when understanding attachment and attachment-related forms of communication.

1.3. Signalling can be manipulated. Honest signals and Machiavellian deceit

Adding one further layer of complexity, signallers can, in turn, foresee a given recipient's reaction to a signal and take advantage of it. This leads to a third nuance: whether signals are univocal— the degree to which they can be reliably interpreted through common past behaviour —and therefore how 'honest' they are.

Animals struggle to avoid being outsmarted and deceived, both intra and inter-specifically. The deceptive capacities of different species have often been studied as expressions of animal´s Machiavellian intelligence (Harms, 2004, Whiten & Byrne, 1988). In this regard, honest signals and the handicap principle are useful concepts. First proposed by Zahavi (1977), the handicap principle proposes that signalling between organisms is kept honest by an intrinsic relationship between the 'meaning' of a complete signal and its production cost. Signals are selected by evolution because they 'handicap' the organism so that their costliness— the signaller's investment in the signals —makes them reliable.

Let us illustrate these concepts with an example that will prove relevant in further chapters. The crying of a helpless human infant is an honest signal. Its acoustical features are the result of significant vocal effort. Such an effort is recognized as a signal of investment by surrounding adults— potential caregivers —, for whom ignoring the signal then becomes almost impossible. A very different case is that of an older child who has learned that crying can be used in Machiavellian terms: pretending to cry will often augment their degree of influence on the caregiver's behaviour. However, mock crying is less costly than genuine crying, and a caregiver can perfectly tell the difference between the two. As a result, the adult may ignore a signal that is not honest in Zahavian terms. As will be discussed in the next chapters, something similar applies to the IDReg, the acoustic qualities of which have evolved so as to constitute an honest signal of caregiving investment.

2. Avian, mammal and non-human primate vocal communication

After establishing and critically addressing some basic principles of animal communication that will prove to be useful in further sections, it is now possible to compare animal communicative dynamics that exhibit close correspondences with human communication. At the same time, it is necessary to narrow down our approach to animal communication into our main interest: vocal communication. Vocal communication is central because of many reasons but let us for now say that it enables "out of sight", distal communication, and that this will play an essential role in considering human infants' early interest in music. As so far noted, animal signals can be used in different scenarios, such as outsmarting prey or communicating with members of the same species while avoiding predators. Nevertheless, the majority of animal signals— and the most complex ones for certain —happen in the context of within-species communication: more specifically, in the modulation of socio-affective dynamics.

2.1. Birds and mammals

Within the animal kingdom, mammals and birds provide instances of precursor systems linking with human aural communication. Mammals and birds both vocalize (albeit at widely varying levels of complexity) in social contexts, these capacities representing shared derived characteristics (synapomorphies) from their common ancestor: reptiles. The emergent complexity of language, Sinha (2004) suggests, has in the course of evolution co-opted or captured a suite of cognitive capacities that are uniquely developed (but not themselves unique) in humans, and constitute an elaboration of capacities already present in our evolutionary ancestors. Many aspects of human language are built on foundations shared with other animals, including, uncontroversially, most aspects of the vocal production and hearing apparatus (Fitch, 2004). For instance, Owren (2011) states that both source-filter production and indexical cueing are deeply rooted in the phylogeny of human vocalization, which becomes clear when bearing in mind both mammalian (in general) and primate (in particular) pasts. Furthermore, he suggests that *'vocalization is as fundamental to being a mammal as having three middle-ear bones or being homeothermic.'* (Owren, 2011, p. 25). However, while vocalization might be widespread amid mammals, vocal learning is a much rarer occurrence. Beyond our species, the phenomenon has only been irrefutably demonstrated in whales, dolphins (Janik, 1997), and bats (Knörnschild et al., 2009).

Although reptiles do have the most basic structures of the limbic system (leaving aside cortical elements like the hippocampus or the cingulate cortex), their rather dichotomous emotional activation (essentially related to fight/flight behaviour) stands rather aside from the richer and more nuanced socio-affectivity afforded by the neocortex in mammals and the pallium in birds (Ahumada-Galleguillos, Fernández, Marin, Letelier, & Mpodozis, 2015). Despite the phylogenetic distance between birds and mammals, the vast majority of avian and mammalian pallial neural territories appear to share a high degree of homologous correspondence. However, the complex interconnectivity the pallium affords manifests in socio-affective life only in some cases. Rooks, corvids, parrots and *passerines* in general (and possibly other large-brained birds) have evolved similar socio-cognitive abilities to primates, while others have not (e.g. geese and albatrosses). These socio-cognitive abilities can underpin pair-bonding or 'relationship intelligence', rather than larger social network relationships. By adding the close, long-lasting infant caring that characterises mammals and socially-intelligent birds (Emery et al., 2007), a correlation between the use of complex sound signals and social bonding becomes evident.

Birds that show collective breeding behaviour and also exhibit learned acoustic signals that are kin or family-specific become an example once again. Songbird has been particularly well studied by behavioural scientists because it is reliably repeated, enabling the experimental control of vocal changes at the population level and across generations in the way of a veritable 'cultural' evolution (Lipkind & Tchernichovski, 2011). The mastering of such acoustic signals requires an early, flexible vocalization process that infants seem to spontaneously exploit (Sharp & Hatchwell, 2006). Such a signal specificity has been referred to as 'mother tongues', and has been acknowledged in many precocial birds (grouses, ducks, etc.) (Matthews, 1969) and mammals (Hoogland, 1983) in the form of food and alarm calls.

An illustrative example of this correlation is avian use of contact calls. In contrast to the longer and more striking bird songs, contact calls are defined as short, simple vocalizations produced by both sexes to provide individuals with information about the location and identity of the caller, and are thus important in mediating reproductive and social interactions (Collias, 2000; Farabaugh & Dooling, 1996; Seddon, Tobias, & Alvarez, 2002; Wanker, Apcin, Jennerjahn, & Waibel, 1998). Although they seem to be ubiquitous among birds (including oscines —"true" songbirds —and non-oscines), different species show different levels of complexity in their use. An interesting case is the (oscine) Long-tailed Tit. Within the vocal repertoire of this species (five call types) one of their contact calls is highly individual-specific and provides cues for kin recognition (Sharp & Hatchwell, 2006). Indeed, some of the acoustic features of the 'churr' (a short-range contact call) are collectively fine-tuned during a family's nesting. As a consequence, variation in the production of churrs by adult non-siblings is significantly greater than that in siblings, thus acting as a cue for kin recognition. A remarkable aspect of this phenomenon is that the family- and individual-specific acoustic features of this contact call seem to develop through learning during infancy (Stuart, Andrew, Matthew, & Ben, 2005). Indeed, although evidence for vocal learning exists in aquatic mammals (seals and cetaceans), their complexity seems to be clearly overshadowed by passerine birds (Fitch, 2000). All oscines studied have shown song learning to a certain extent, requiring exposure to conspecific song during infancy in order to develop socially-successful ‘songs’. More specific subgroups of species like that to which mockingbirds belong, show an impressive degree of flexibility and proficiency in the learning of new sounds, imitating the songs of other bird species, along with environmental sounds like crickets and even car alarms (Ibid).

As previously mentioned, cetaceans also evidence learned vocalizations which can function as an indicator of group membership. Researchers have characterized some of these as ‘dialects’, which can be shared by a population, a social group or extended family. Such is the case of ‘signature whistles’ seen in bottlenosed dolphins, which consist of specific frequency contours that seem to enable individual recognition by voice alone (Sayigh, Tyack, Wells, & Scott, 1990). Furthermore, males adopt their mother’s signature whistle before emigrating to distant waters, which would allow brothers who have never met to recognize each other as kin. Similar dialects that signal group membership or kinship in killer whales have also been reported (Ford, 1991). Like chimpanzees and humans, these mammals generate social groups characterized by within-group cooperation and between-groups competition, thus relying on indicators of group membership, vocal or otherwise.

Thus, general avian and mammal communication may work through acoustic or auditory signals. These signals may show variations, parents being responsible for exposing infants to them, and infant signallers being responsible for their learning and performance (e.g. infant long-tailed tits learn and slightly tweak their family-specific version of a churr).

Interesting advancements have been made regarding the constraints that guide the ontogenetic evolution of songbird and the extent to which it applies to human vocal learning. Comparative study of the zebra finch and Bengalese finch revealed a common, stepwise pattern in the acquisition of vocal transitions that also applies to 9-28 months-old children (Lipkind et al., 2013). In short, newly learned syllables tend to be repeated (i.e. the same syllable over and again) when first acquired, to be only gradually concatenated to other available syllables.

In the same vein of vocal variation constraints, birds reared in captivity and exposed to particular, artificial songs show nevertheless a tendency to converge on 'universal', 'wild-type' song structure in their final outcomes (James & Sakata, 2017). In other words, a family-specific churr is still clearly recognizable as a churr. Its general structure has slowly evolved to become what it is and birds have long evolved to recognize it as such, thus qualifying as animal communication in Smith and Harper's (2003) canonical terms. Still, it remains remarkable that a churr or a dolphin's signature whistle convey affiliation not simply through indexical information (e.g. signaller's arousal level or size), but through complex, socially learned variations of a signal. Some of these socially-led vocal learning dynamics are also exploited by human infant-carer dyads, as will be discussed when examining the IDReg in chapter 4.

2.2. Non-Human Primates

As previously hinted, the number of scientifically acknowledged continuities between the communicative capacities of human and NHP constantly increases. As in the case of previous topics covered in this work, interest in these continuities— and the remaining discontinuities —is partially fuelled by the agendas of ethology and compared psychology. However, given the evolutionary proximity of humans and NHP, such an interest is more heavily complemented by linguistics, psycholinguistics cognitive science, and anthropology in this case. Because of these reasons and considering current purposes— a focus on socio-affectivity —an exhaustive account of NHP communication is not necessary, nor would it be the most convenient way to proceed. Instead, it will be more fruitful to address the continuities— and discontinuities — between NHP and human communication shown to be relevant in contemporary research. Proceeding in this manner will allow the present argument to smoothly approach the next section: human communication.

In the introduction to their book "Mind the Gap" (Kappeler Silk, Burkart, & van Schaik, 2010), Kappeler and Silk are quite clear when addressing human and non-human primate similarities: following the cultural and social complexity that apes show, the gap that separates humans from its closest relatives is a matter of degree, rather than of fundamental innovations. Moreover, the existing innovations are primarily behavioural and cognitive, rather than genetic. Whatever the selective pressures that prompted such changes, Cheney & Seyfarth (2010) conclude, the complex suite of skills that we call human speech build upon mental computations that had their origins and foundations in social interactions. As Kappeler and collaborators (2010) point out, in order to understand what makes us different from the rest of primates it is necessary to look at more than communication alone. Family and social organization, politics and power, intergroup relationships and foundations of cooperation are some of the intimately related matters. Nevertheless, for current purposes let us focus on communication, heuristically breaking it down into its gestural and vocal domains.

Primate communication— especially its more complex aspects —remains mainly gestural. NHPs show a range of voluntarily controlled signals, which has prompted theories proposing a gestural origin of human language (Hewes et al., 1973). Other researchers, while agreeing on locating the antecedents of language in primate cognitive abilities, deem the emergence of language from any earlier primate communication system implausible, including both animal use of tone of voice and gestures. Their main argument is that gestures and tone of voice (our 'surviving primate communication system') remain up to present sharply distinct from language despite closely cohabiting with it (Burling et al., 1993). In any case, when examining NHP communication between conspecifics, researchers have highlighted the flexible usage of gestures, considering evidence for intentional signalling and the potential to learn and generate novel gestures (Call & Tomasello, 2008). This does not mean, however, that new gestures acquire shared social value like new words do. Still, it is through gestures, not vocalizations, that attempts to teach apes human language systems, such as American Sign Language, have succeeded (Gardner & Gardner, 1969; Hayes, 2008). Finally, the interest in NHP gesture has been fuelled by the discovery of mirror neurons and the apparent link between manual gestures and homologous areas of language production in the monkey brain (Arbib, 2005).

When it comes to complexity in the vocal domain, humans seem to belong to a 'disjoint' group that includes birds and aquatic mammals, but skips our closest relatives: NHP (Fitch, 2000). NHPs usually have a small repertoire of calls that show relatively little modification in their acoustic features during development and remain largely unaffected by variations in auditory experience or rearing. For example, male baboons produce loud two-syllable alarm 'wahoos' when they encounter lions or leopards. Alarm wahoos are acoustically similar to the contest wahoos that males give during competitive contests with other males, but the two types of wahoo differ according to a number of acoustic measures (Hammerschmidt & Fischer, 2008). NHP calls thus can refer to external objects and events, arguably constituting a precursor to human referential abilities (Seyfarth, Cheney, & Marler, 1980; Seyfarth, Cheney, & Marler, 1980). Unlike predator alarm calls, which depend in a fairly simple way on the type of predator or the degree of danger, the vocalizations given by animals during social interactions are elicited by a more complex array of factors that may include both the immediate social context and the history of interactions between the particular individuals involved (Cheney & Seyfarth, 2010). NHPs do not seem to be able to generate novel vocalizations in the way they do with gestures, however, the vocalizations of some of them can be combined into sequences that are produced in response to specific context (Ouattara, A. Lemasson, & K. Zuberbühler, 2009a), indicating that simple rule-based combinations exist in primate vocal behaviour (Ouattara, Lemasson, & Zuberbühler, 2009b).

Along these lines, it has been argued that NHP vocalizations can be perceived as discrete signs, and thus have the potential to be combinatorial— a cornerstone of language. Vocal theories emphasize that primate call perception is complex and the ability to assign meaning to calls and call combinations is highly flexible (Arnold & Zuberbühler, 2008; Seyfarth, Cheney, & Bergman, 2005). For instance, acoustic intergradation in many NHP vocalizations may be caused by gradation in the caller's arousal or emotional state. Analyses of baboons' contact and alarm barks, contest and alarm wahoos, and move and infant grunts all suggest that acoustic variation is consistent with variation in the caller's emotions (Fischer, Metz, Cheney, & Seyfarth, 2001; Rendall, 2003). By relying on the signaller/receiver debate previously reviewed, it is possible to highlight that it is the receiver— and not the signaller —that considers the unintended elements (arousal level) of a signals and uses it as indexical cues to be added to the intended element (the call itself). It is thus in terms of call perception rather than call production that the greatest similarities with human language seem to lie (Seyfarth et al., 2005).

By now, it is probably evident to the reader that there is an integrated use of the different communicative modalities of NHPs, and that the gestural and vocal dimensions complement one another (a point not always considered, see Slocombe, Waller, & Liebal, 2011). While exclusive focus on vocal or gestural production reveals clear differences between humans and most species, a broader examination of call usage and (especially) perception delivers a more complex scenario: while it may be true that NHPs have only a small repertoire of acoustically-fixed vocalizations, because calls are individually distinctive and each call type is predictably linked to a particular social context, this limited call repertoire can, nonetheless, provide listeners with a highly modifiable, cognitively rich, and— potentially —open-ended set of meanings, allowing them to construct “narratives” of unseen events (Cheney & Seyfarth, 2010).

Thus, unlike in the case of birds and other mammals, NHP communication seems to consistently go beyond the dynamics of Smith and Harper’s (2003) definition of animal communication; learning a completely new gestural sign taught by a human is no doubt very different from introducing variations to a long-evolved gestural or vocal signal. The particular combination of signals with other signals according to a particular context gives them communicative value other than the one they evolved for.

Still, as Cheney and Seyfarth (ibid.) point out, although NHPs and other animals seem capable of thinking, as it were, in rudimentary 'sentences’, this ability does not motivate them to actually speak in sentences, their knowledge remaining largely ‘private’. NHPs lack of ability to generate new words or a lexical syntax partially explain this. However, Cheney and Seyfarth attribute this gap, mainly, to NHPs’ lack of another key element: a theory of mind. Although chimpanzees do seem to assume another individual’s role in a cooperative task and even recognize some intentional gestures such as pointing, results of tests for a theory of mind are mixed (Povinelli, 1996)**.** While there is evidence from different experimental paradigms that chimpanzees understand the goals and intentions, perception and some of the knowledge of others, there is currently no evidence that chimpanzees understand false beliefs. In sum, to date, chimpanzees seem to understand others in terms of perception–goal psychological dynamics, but not quite in the full-fledged belief–desire psychology which adult human interactions rely on (Call & Tomasello, 2008; Meunier, 2016).

Indeed, while animals are concerned with their own goals and knowledge, Cheney and Seyfarth argue, young children are readily and constantly concerned with making their thoughts and knowledge publicly available, selectively exposing their communicative behaviour towards potential witnesses. On the other hand, NHP and other animals fail to distinguish between what they and others know and therefore cannot recognize, for instance, that an ignorant individual might need to have an event explained. As a consequence, the 'linguistic revolution' would have occurred when human ancestors began to express this tacit knowledge and to use their cognitive skills in speaking as well as listening.

What missing element is then preventing NHPs from 'speaking'? What conditions could have and continue to motivate human beings to share their psychological states from so early on? Answering these questions will imply partially leaving behind animal communication and moving closer to our main subject: the role of human communication in bonding. The word 'partial' reflects the fact that it seems to be mainly social practices common to birds, mammals and NHPs that provided— and continue to provide —the conditions for human communicative complexity. Along with this, the second section of the first chapter will explore the particularities of human anatomy and interaction that to this date provide the strongest explanations for the emergence of language as an alternative to animal communication. Ironically, it will not be in language *per se* that the most relevant link between human communication and bonding will be found.

Part II. Human communication: a multi-layered ‘system of systems’

1. Human communication is *and* is not animal communication

So far, I have built an idea of how animal communication has been generally conceived, structured, and problematized. Animal communication mainly relies on signals that a) influence the behaviour of other organisms, b) have evolved to that effect, and c) are effective because the receiver’s response has evolved along with the signal (Smith & Harper, 2003). Such an evolutionary-led feedback dynamic is central. Although the idea can be defended that some kind of ‘information’ is transferred in animal signalling processes, the impossibility of accessing its ‘content’, added to an unsystematic theoretical framework, makes influence a simpler and often more reliable construct. In any case, animal signalling is greatly complemented and enriched by the intelligence of the signaller and the receiver. For instance, the signaller may often outsmart and thus ‘deceive’ the receiver in a ‘Machiavellian’ manner. Conversely, the receiver may extract additional complexity (e.g. consider indexical cues, combine signals) thus exceeding the purpose of the signaller’s communicative complexity. Finally, layers of complexity can be intentionally added, as some avian and mammal infants do by learning group-specific signals, mainly for socio-affective ends.

The next questions in this argument then are: where do the closest similarities between human and non-human animal communication lie? To what extent does human communication rely on the kind of animal signals and biological communication so far presented? What factors may have contributed to separating human communication from NHPs? And, just as importantly, what role does all of this play in human bonding or music?

In order to answer these questions, it will be necessary to briefly explain the multi-layered nature of human communication, and explicitly distinguish language from the rest of such layers. Having done so, it will not require a long discussion to posit that language is definitively not animal communication, while some aspects of speech and general human non-verbal communication— notably, prosody —largely remain so. Nevertheless, as the reader must have already inferred from the subsection on NHP communication, most of the study of human communication and its (dis)continuities with the case other animals focuses on language. Therefore, although language *per se* will not be central to this argument, the study of its evolution will. As already mentioned, the existing evolutionary innovations in human communication are primarily behavioural and cognitive, and have their foundations in social interactions. Thus, in accounting for the human social dynamics that led to language, the thread will come down to human infant communication, when language has not been learned yet.

1.1. Language is not animal communication

In this thesis I will refer to 'language' as a communicative system learned by humans that relies on distinct, culture-specific signs: words. Put simply, 'language' will refer to lexical language. Needless to say, there are many approaches and definitions that have been used to account for language as a general phenomenon, stressing its different aspects. Amongst them, the one that stands furthest away from the rest of animal communication is its status as a formal system of signs, governed by grammatical rules of combination that convey meaning. This definition, originally put forward by Ferdinand de Saussure and the structuralist school, describes language as a closed structural system consisting of rules that relate signs to meaning (Campbell, 2001). Indeed, when analysed as a system, language works as a referential, combinatorial structure capable of enormous flexibility and complexity, linguistic meaning residing in the logical relations between propositions and reality. In other words, rather than meaning residing in one or another particular word, it is the mutual relationship between words and context that enables its emergence. As a result, humans can produce a potentially infinite number of novel— new, non-established —

structures, a property known as productivity (Trask, 2007). Cultural regulation of the use of these possibilities in everyday communication has spontaneously occurred in all known human populations, and syntax is the resulting set of rules, principles, and processes that govern the structure of sentences in a given language (specifically word order and punctuation). As previously discussed, although some NHPs can spontaneously combine signals from a fellow ape, these combinations have not been shown to be shared or established at a group level. Another key human innovation is the property of displacement: the capability of referring to things that are not immediately present, be it spatially or temporally (Ibid). Furthermore, language can refer to things that no one has ever seen, like dragons, or to abstractions or generalizations of our experience, like concepts. Also unique in the animal kingdom is the property of recursivity: a noun phrase containing another noun phrase or a clause containing another clause. Human language is also the only known natural communication system whose adaptability may be referred to as modality-independent. This means that it can be used not only for communication through one channel or medium, but through several. For example, speech— mainly —uses the aural modality, whereas sign languages and writing rely more heavily in the visual modality, and braille writing uses the tactile modality.

Because of these differences, there is consensus among scholars when stating that the referential/combinatorial dimension of language definitively does not work in the manner of animal signals, and stands as a uniquely complex system that clearly separates human communication from the rest of the animals, including NHPs. Such a complexity largely allowed the emergence of civilization, philosophy, and written language, and therefore attracted the attention of formal logicians, formal theorists of grammar, and applied computational linguists all over the 20th century. Lexical ('word') language remains a cornerstone of our species and is the basis of informatics and the computational sciences that characterize contemporary ‘western’ societies.

1.2. Human prosody often works as animal communication

Nonetheless, as mentioned in the beginning of the second section of this chapter, human communication is a multi-layered phenomenon, broader and richer than language alone. As Levinson and Holler (2014) point out, the gap between animal and human communication systems reported by scholars can be largely explained by their focus on the presence or absence of language as a complex expressive system built on speech. Such a focus largely bypasses the fact that language normally occurs embedded within an interactional exchange of multi-modal signals. A subset of such signals constitutes what Stephen Levinson (2006) called the ‘human interaction engine’, consisting of the parallel and integrated use of speech, gesture, face-to-face multiple parallel articulators, mutual gaze and sustained exchange of fast and complex turn taking —basically, all of the upper half of the body participates, something not possible before bipedalism. The sum of these elements constitutes human ‘interactional intelligence’, the necessary niche and core ecology of language (Ibid). As such, human communication is better described as a ‘system of systems’ (Levinson & Holler, 2014); rather than vocal, facial and bodily gestures scaffolding the evolution and ontogenetic development of speech, these two modalities/systems are different, complementary components of one and the same human communicative meta-system. Such a complementarity remains flexible, enabling human beings to shift the burden from words to other, non-verbal forms of communication as required by communicative needs.

As true as it may be that when compared to language (and its unprecedented communicative properties) the rest of human communicative systems are not as effective in referring to distant or abstract matters, such a disparity does not apply when it comes to handling real-time communicative situations. As explained in the first part of this chapter, indexical cues can be extracted by a communicative recipient and added to whatever the signaller may intend to communicate in the first place. In other words, intelligent-enough animals are able to obtain 'information' about the physical (i.e. size) and psychological (i.e. arousal level) state of an animal while it communicates, and humans are no exception to this.

The main construct to refer to all of the indexical 'information' that humans are able to extract from the physical act of communication is prosody. Prosody can reflect various features of the speaker, as well as both linguistic and paralinguistic aspects of the utterance: the emotional state of the speaker; the form of the utterance (statement, question, or command); the presence of irony or sarcasm; emphasis, contrast, and focus; and other elements of language that may or may not be encoded by grammar or by choice of vocabulary in addition to prosodic markers (Hardcastle, Laver, & Gibbon, 2010). Furthermore, prosody can not only provide 'information' about a particular individual's physical or psychological state, but about their position regarding the current communicative situation and the recipient(s) of the message. Indeed, attitudinal prosody (Mitchell & Ross, 2013) can reveal an individual's approach towards an event, person or object, and convey relevant attitudes such as confidence, persuasion, sarcasm and superiority.

Although prosody is a multimodal phenomenon, its vocal component is central. The importance of vocal prosody varies contingently, depending on the situation. Prosodic auditory signals remain important on a daily basis and can be particularly crucial at times as they allow, for instance, affective communication when the recipient cannot see the sender across a distance or at night (Sauter, Eisner, Ekman, & Scott, 2010). Further relevance to the vocal domain is added by the fact that, unlike language (which is highly specific at various levels such as culture, region, group, etc.), vocal prosodic cues— particularly emotional ones —can be effective cross-culturally. In other words, listeners can infer a number of affective states from emotionally inflected speech, across cultural boundaries (Bryant & Barrett, 2008; Pell, Monetta, Paulmann, & Kotz, 2009; Scherer, Banse, & Wallbott, 2001). Non-verbal cues other than the prosody that accompanies speech, such as screams and laugh, are also effective across considerably different cultural groups. Vocalizations communicating the so-called 'basic emotions' (anger, disgust, fear, joy, sadness, and surprise) are bi-directionally recognized (Sauter, Eisner, Ekman & Scott, 2010). Interestingly, it is the vocal correlates of primarily negative emotions that can be recognized across cultures, while positive emotions seem to be communicated through more culture-specific signals.

Indeed, prosody is not only effective across languages, it can also be effective on organisms that do not exhibit linguistic capacities yet, such as human infants. The relevance of vocal prosody varies throughout the lifespan, as infants are particularly sensitive to vocal cues from the very beginning of life, when their visual system is still relatively immature (Mehler, Bertoncini, Barrière, & Jassik-Gerschenfeld, 1978). During such a period, the value of vocal prosody is highlighted, as it constitutes a privileged channel of interaction, and a form of biological communication— as defined in the first part of this chapter. Given such a pivotal role, cues embedded in the caregiver's voice are exaggerated and made more explicit, a quality that led to one of the first characterizations of the IDReg (e.g. Ferguson, 1977). As will be discussed in detail in the next chapter, although infants' visual system eventually matures, their sensitivity to vocal prosody remains crucially important, and so will the mentioned register.

1.3. Non-verbal cues: honest signals in human communication

Prosody thus provides concrete and familiar insight into the multi-layered quality of human communication. Always in parallel to the logical-referential exchange that language allows, human beings consciously or subconsciously extract as much 'information' as possible from interlocutors, just like any other intelligent animal. Accordingly, the interpretation of prosodic cues does not normally work through the complex combinatorial dynamic of language (although, exceptionally, it can), but through those that most mammals and NHPs rely on.

Precisely because of their interrelation with rather spontaneous, non-reflective psychological processes, prosody and non-verbal cues, far from an evolutionary vestige, remain highly functional and relevant in human social life and have been proposed to occasionally still work as an honest signal (Pentland & Heibeck, 2008). As explained in the first part of this chapter (section 1.3), Zahavi's (1977) handicap principle states that animal signalling is kept 'honest' by an intrinsic relationship between the evolved 'meaning' of a complete signal and its production cost. Accordingly, attempts at accurately faking an emotion or the precise timing and expression of spontaneous attitudes require considerable effort, and are more often than not unsuccessful. Indeed, we, humans, are socially trained to distinguish the genuine from the faked when it comes to smiles, yawns, tears and tantrums. As a result, faking emotions and other normally spontaneous behaviours is comparatively much harder than assembling the verbal component of a lie[1]. Accordingly, honest signals are naturally convincing and have thus been reported to induce empathy, mood contagion, mutual trust and group cohesion (Pentland, & Heibeck, 2008). Hence, whenever careful logical appraisal is not possible (i.e. lack of factual information, time constraints), accurate human judgments are made on the basis of prosodic cues (Pentland, 2010). The fact that

[1] The possibility of Machiavellian deceit is not directly problematized in Pentland's work. A skilful actress or a perverse psychopath may well generate 'fake honest signals' that manage to deceive the receiver. As Lachmann, Szamado and Bergstrom (2001) suggest, since language facilitates the spread of reputations by allowing individuals to share information about others' reliability, more important is the implication of an interlocutor's claim rather than honesty per se in a given moment, largely dissociating meaning and cost. Effectively, the aforementioned authors recognise that 'conventional' signals— words —will be used when communicating about (a) coincident interests or (b) verifiable aspects of conflicting interests; 'costly' honest signals will be used otherwise. Therefore, when encountering strangers, in the case of sexual behaviour like courtship, and most likely in situations of actual peril or violence, the most persuasive signals will remain genuinely costly (Fessler & Gervais, 2010).

prosodic honest signals work through influence rather than conscious, critical reasoning, is another aspect that links human communication to the dynamics of general animal communication. Thus, honest signals[2] such as spontaneous (non-faked) emotional and attitudinal prosody impact our everyday intimate life, as well as public activities including negotiation, group decision making and group management. They highlight the pervasive role of influence— as opposed to 'information' —in all communication, including daily human interaction.

2. The specifics of human communication evolved largely as a response to the specifics of human caregiving

By now, I have delved into how animal communication has been generally conceived and problematized, and to what extent human communication work through the dynamics of animal communication. I have established that human communication is better described as a multi-layered, multimodal 'system of systems'. I have also established that language— the exclusively-human layer —works as a referential, combinatorial system whose properties even our closest and smartest NHP relatives cannot collectively exploit, thus stepping aside from the rest of animal communicative systems. At the same time, non-verbal layers of human communication such as prosody— particularly in its affective component —do work by means of evolved feedbacks systems, and therefore often show no significant difference from animal signals.

[2] The reader may have well spotted that in a considerable number of instances prosodic cues are in fact— forgive the repetition —cues and not biological signals, as previously defined by Smith & Harper (2003), in the sense that they have not necessarily evolved to generate a particular impact. Nevertheless, spontaneous affective prosody, especially the one that conveys the so-called 'basic emotions', will more often than not exert the effect these emotions evolved for.

As discussed earlier in section 2.1, it is in the intimate social environment of infancy, and in all likelihood for social reasons, that some birds and cetaceans evidence fine-grained vocal learning. Although it has not been completely demonstrated, the evolution of capacities for sociality and complex vocal communication do seem to be intertwined. In the remainder of the present chapter, a brief review of the evolution of language will be offered, with a focus on the peculiarities of human social (in general) and socio-affective (in particular) dynamics. More specifically, it will be shown that, as in the case of other animals, the combined selective pressures of gregarious life and attachment led to the emergence of cooperation, cooperative breeding and different kinds of kin-specific communication like mother tongues, infant-directed specialized communication, and different forms of infant-carer acoustic feedback. As we will see, the evolution of bipedalism in human descent introduced changes in cooperative breeding and kin-specific communication that largely account for the communicative 'gap' that separates us from other animals.

Language is the culmination of a slow and progressive process of increasing diversification of social bonding mechanisms based on natural forms of communication, the principal selective advantage of which was social rather than environmental or technical (Dunbar, 1996). In this view, language evolved to supplement and, eventually, largely replace grooming as the principal mechanism for social bonding within the later hominid lineage. This, as it would have maximized efficiency in social time budgeting by allowing multiple (as opposed to one-on-one) hands-free engagement, and larger 'information' retrieval (Dunbar, 2004a). Such an approach stresses communication within large social groups, 'information' exchange concerning the state of the social networks such groups generate, and mechanisms for controlling them (Dunbar, 2004b).

On the other hand, several authors locate the main adaptive value of vocalization for in-group adaptation in the dynamics not of large groups, but in those of the closest interpersonal bonds: parent-offspring communication. These theories examine social dynamics such as cooperation and cooperative breeding, and infant-carer[3] feedback.

Once again, the reason for examining these communicative phenomena is that they stand at the intersection of all this thesis's interests at once; they comprise cases of contemporary human communication that rely on mechanisms rooted in our animal heritage, while at the same time enabling interpersonal bonding. Among them, infant-directed speech, a form of 'attachment vocalization', will constitute the main subject of interest in this work as it additionally relates to the only pending element of this introduction: music.

As it will be explained in the next sections, none of these social phenomena (cooperative breeding, mother tongues, and infant-carer feedback) *per se* is exclusive to human beings. It is rather the degree of refinement they acquired in human contexts that in all likelihood led to language, and they therefore are worth scrutinizing. The question then remains: what let to such a degree of refinement? The common factor is bipedalism. The anatomical correlates of this evolutionary shift in locomotion had tremendous impact in human gestation and, by implication, in human development and breeding. More specifically, it led to human infants being born in a state of helplessness unprecedented in the primate order— secondary altriciality — and to concomitant breeding efforts from their caregivers. Although we will see that the importance of bipedalism and especially of secondary altriciality for human communication and bonding has been increasingly acknowledged, the subject can— and will, in this work— be further explored. Given its importance, let us now dedicate secondary altriciality a section so as to better understand its implications for human cooperative breeding, mother tongues, and infant-carer feedback.

[3] Although the terms 'parent' or even 'mother' would apply for most cases, I will use term 'carer' since it encompasses the first two plus all potential alloparents (uncles, siblings, etc. See Hrdy, 2009). As will be presented in the next chapter dedicated to imprinting, what the infant might identify as the carer need not necessarily to be the biological mother, or even a member of the same species. For the same reason, I will also prefer the term 'infant directed speech' over 'motherese'.

2.1.2.2. The impact of secondary altriciality on cooperative breeding

As discussed in the beginning of the present chapter (Part I, section 1.3), within the broad scope of animal communication and signalling behaviour, the competitive economics of the exploitation of perception relies on resources such as the handicap principle, Machiavellian deceit, and other forms of competition and depredation. In contrast, the emergence of gregarious life and socio-affectivity led to an alternative that allowed radically different communication dynamics without contradicting the principles of evolution: cooperation.

Cooperative breeding can manifest in various group structures. For instance, in breeding pairs, the male mates exclusively with one female, forming a long-term bond and combining efforts to raise offspring together. In polygynandry groups, like Acorn Woodpeckers, an infant can be taken care of by multiple breeding males and females as well as collaborators that are the adult offspring of some (but not all) of the breeders in the group (Haydock, Koenig, & Stanback, 2001). This is the case of most of our closest NHP relatives, as well as that of our species. Cooperative breeding dynamics are indeed present in NHP, and have been argued to partially underlie cooperation in human societies (Silk & Boyd, 2010). Nevertheless, the level of physiological maturity at birth in NHP and humans is crucially different. Let us briefly explain this maturity gap, and then proceed to discuss its significant consequences.

Among bird and mammal ontogenies, different species display two patterns of growth and development, classified either as precocial or altricial. Altricial species are those whose newborns are relatively immobile, are not able to obtain food on their own, and must be cared for by adults. Conversely, precocial species are those in which the young are relatively mature and mobile from the moment of birth or hatching (Starck, 1998). The two categories designate the ends of a continuum rather than purely dichotomous alternatives, thus intermediate cases can often be observed.

In a contrary direction to the general primate trend, a further degree of altriciality— secondary altriciality —emerged in humans (and most probably in our immediate ancestors, see Bogin, 1999) in tandem with pelvic accommodation to the parallel development of both bipedalism and progressive encephalization (Buck, 2011). Secondary altriciality is defined as '*the infant emerging in a state of helplessness due to having only a third of the brain size relative to mature adults*' (Ruff & Walker, 1993, p.277). The subsequent post-birth rate of brain growth is so substantial that anthropologists have referred to this period as 'exterogestation' or 'external gestation' (Rosenberg & Trevathan, 1995).

As a result, the unprecedented helplessness of human infants implied— and still implies —a significantly higher degree of co-dependence amongst group members. Indeed, human mothers rely considerably more in other group members than our closest NHP relatives do (Chang, 2013). For instance, orangutan, chimpanzee and gorilla mothers do not allow others to hold their infants until these reach 3.5–6 months of age, while human mothers do so almost immediately following birth (Hrdy, 2001). Regarding such differences, Chang observes, humans are much more similar to primate species that cooperatively breed, namely marmosets and tamarins. Crucially, in the highly cooperative breeding dynamics that secondary altriciality imposes, the care and protection of the young becomes a bonding mechanism for the members of a group, breaking the discrimination that many species keep between the distress sound of their own offspring and that of others, and turning kin and kith into potential 'alloparents' (Hrdy, 2001). Such a widening of the caregiver spectrum will be relevant for the present thesis. As will be discussed (Chapter 3, section 2.2.2.1), not only parents but virtually any member of society engages to some degree in caregiving-related behaviour— including IDSp —when interacting with children. Accordingly, the latter will also to some degree grow an expectation to receiving such kind of behaviour.

In the same line, Burkart, Hrdy and Van Schaik (2009) argue that while chimpanzees and probably all great apes exhibit many of the important cognitive preconditions from which human mental capacities evolved, what they lack are the psychological preconditions. The authors propose that the two components merged in humans: the cognitive component inherited from ape ancestors and the motivational components that cooperative breeding entails. Van Schaik and Burkart further suggest that cooperative-breeding would have largely afforded the emergence of a declarative communication systems such as language (van Schaik & Burkart, 2010). This idea is supported by Gärdenfors, who also suggest that cooperation precipitates language, in the sense that it provides the motivation for the sharing of cognitive and affective aspects of mental life (Gärdenfors, 2004). Gärdenfors' argument provides insight into what Cheney & Seyfarth's (2010) consider to be the crucial factor behind the communicative gap between humans and NHPs. As discussed, NHP and other animals fail to distinguish between what they and others know. The augmented need for cooperation that secondary altriciality imposes would have at least partially motivated human ancestors to express their tacit knowledge and to use their cognitive skills in speaking as well as listening. The same case can be made for Hrdy's (2001) alloparenting *theory.* Instead of depending on the exclusive dedication of their mothers, human infants had to additionally monitor and engage multiple caretakers. Thus, while other NHPs may attain a rudimentary theory of mind, it would have been in the context of cooperative rearing that relevant potentials for 'mind reading' would have become more overtly expressed, and thus naturally selected as social mechanisms. Over generations, Hrdy argues, collectively-bred infants who were used to inter-subjective engagement would have been best looked after and fed, leading to a selection process that favoured human capacities for intersubjective engagement.

Accordingly, among the different forms of cooperation mentioned by van Schaik and Burkart (2010), cooperative breeding would stand out as the main aspect responsible for socio-cognitive traits such as pedagogy, extensive cumulative culture and cultural norms, intensive and nearly indiscriminate within-group cooperation, and morality.

2.1.2.4. Infant-carer acoustic feedback

On the subject of secondary altriciality and its consequences for parenting, it has been proposed that at some point of the australopithecine/early Homo transition— during which the emergence of Homo heidelbergensis occurred —maternal pelves that had been modified to accommodate bipedalism became subject to an emerging trend for increasingly large brains, which eventually caused a selective shift toward females that gave birth to relatively undeveloped and helpless neonates (Falk, 1998). Consequently, the ability of babies to cling actively to their mothers was lost in hominins. As a result, the incidence of distal mother-infant gestural communications increased (Tomasello & Camaioni, 1997) and prosodic affective vocalizations became ubiquitous to compensate for the reduction in sustained mother-infant physical contact. This is particularly relevant given that, as discussed, affective prosody works through the dynamics of animal communication. This influence of human anatomy in sociality and communication is utterly central in the present argument and will be approached from different angles in the following chapters.

Since language acquisition is today universally scaffolded onto infant-directed speech (IDSp, or motherese), Falk (2004) further argues that selection for the pre-linguistic vocal substrates supposed to be present in protolanguage occurred after early hominin mothers began engaging in routine affective vocalization toward their infants as a compensatory socio-affective measure. Falk also posits that the prosodic features of protolanguage would find their closest simile in contemporary IDSp. A contrary position to this line of thought is that in spite of its intimate cohabitation with word language, tone of voice remains as part of a distinctively different human gesture-call system (Burling et al., 1993). In any case, the ideas of human physiological prematurity and its consequences for communication are aligned with, and yet different from the previously discussed extended childhood. They are different as they impute greater biological urgency to vocal communication when compared to previous animal examples and bring forward a corresponding psychological experience related to it. Thus, in Falk's thinking sound inherits the socio-affective power of touch in a 'communicating vessels' logic: as touch becomes less available, sound gains equivalent importance. The same can be said about Dunbar´s transition from grooming to language.

At this point the signaller/receiver debate comes into play. It is necessary to consider not only the role of the caregiver as a signaller but also as the receiver of vocal communication. Following Cheney & Seyfarth's (2010) arguments, a crucial point to be made is that the poverty of the signalling repertoire of a physiologically premature infant does not preclude the possibility that adult hearers may still react in multiple, complex ways to it. Precisely, by approaching it from the intimate social regulatory context of caregiving Owings and Zeifman (2004) conceptualize infant crying as a sophisticated form of assessment-management dynamic (and so, without acknowledging it, does Falk in her proposal of vocalization as a compensatory socio-affective measure). According to Owings and Zeifman, for crying to be effective in securing the attention of kin among the rest of the larger group, it must have the power to capitalize on the motivational and emotional systems of the intended receiver as a signal that activates the sympathetic nervous system (Crowe & Zeskind, 1992; Frodi, 1985) and therefore cannot be ignored. Since the adaptive value of crying is perhaps most obvious for the infant at a time when he or she is helpless to meet his or her own needs, crying has been considered the 'acoustical umbilical cord' that ties an infant to its source of sustenance (Ostwald, 1972).

Owings and Zeifman´s proposal draws simultaneously on ethological and psychological ideas. One the one hand, it relies on Lorenz's (1970) ethological research and his concept of imprinting; an important process whereby animals develop the ability to recognize and become attached to companions. The idea of imprinting rests in turn on kin-specific communication such as cooperative breeding and mother tongues, both reviewed in previous sections. On the other hand, the concept of imprinting lead to the development of attachment theory (Bowlby, 1969), according to which the perception and action systems of human infants need to be understood in terms of the properties of the infant's most significant initial companion, typically their mother. Human infant crying can thus be understood as part of the feedback-sensitive ´attachment behaviour´ that serves to maintain the proximity or availability of the attachment figure, having the infant's need for security (or perceived safety) (Bowlby, 1969) as the homeostatic 'set goal' maintained by the attachment system (Owings & Zeifman, 2004).

Thus, it seems as though the development and specialization of human vocal communication— as well as its progression into language —largely find their roots in the mutual accommodation between caregiver and infant to the requirements set by bipedalism on successful breeding. In an assessment-management dynamic, the carer transmutes affective stimulation from the tactile into the acoustic domain, and the infant in turn refines its acoustic signals. In doing so not only the biological, but also the psychological integrity of the infant are secured.

2.2. Closing remarks

Let us recapitulate. In this chapter, we have first delved into how animal communication has been generally conceived and problematized, and to what extent human communication does and does not work in the parameters set by the dynamics of animal communication. I have established that human communication is better described as a multi-layered, multimodal ‘system of systems’. I have also established that language— the exclusively-human layer —works as a referential, combinatorial system whose properties even our closest and smartest NHP relatives cannot collectively grasp, thus stepping aside from the rest of animal communicative systems. At the same time, non-verbal layers of human communication such as prosody— particularly in its affective component —do work by means of evolved feedbacks systems, and therefore present no fundamental differences from animal signals. A general review of the evolution of language focused on the peculiarities of human social (in general) and socio-affective (in particular) dynamics. As in in the case of other animals, the combined selective pressures of gregarious life and attachment led to the emergence of cooperation, cooperative breeding and different kinds of kin-specific communication such as infant-directed specialized communication, and different forms of infant-carer acoustic feedback.

All of these social dynamics point, in one way or another, to infant-carer-centred relationships, where language is not yet an effective medium. In particular, the idea of infant crying and infant-directed vocalization working as corresponding elements of an attachment system is extremely relevant to this thesis’s argument, since it points up a case of human communication that can both be considered a form of biological communication while at the same time standing at the heart of human bonding. Regarding communication, let us remember that whereas signalling or cueing may occasionally fail, biological communication refers exclusively to the successful completion of a signalling act (Scott-Phillips, 2008). Accordingly, on the one hand an infant being distally soothed, appeased or calmed, or engaged in protoconversation (see chapter 4, section 3.2), would imply the successful completion of the biological communication attempted by infant-directed vocalization. On the other hand, same case can be made for an adult feeling an irresistible urge for caregiving as a response to infant crying.

I have by now built a good-enough understanding of animal and human communication and their interrelationship, as well as finding their main relationship to bonding. The moment has thus come to build an equivalently solid understanding of animal and human bonding: attachment theory. As explained in this chapter's introduction, attachment is key to this dissertation, insofar understanding some of its aspects will allow us to properly lay out the motivational aspect of vocal-affective weaning.

CHAPTER 3. A SOURCE OF MOTIVATION: ATTACHMENT THEORY

The second chapter's aim was building a good-enough understanding of animal and human communication as well as their interrelationship. While accomplishing such an aim, literature showed that some of the most important cases of biological communication present in human interaction are intimately related to interpersonal bonding. Given that the latter is also at the centre of the present thesis, this third chapter will be dedicated to building a solid understanding of animal and human bonding. This enquiry will keep in mind biological communication (evolved signalling feedbacks) and general animal communication dynamics (indexical cueing, etc.), and will highlight any possible role they may play. In this manner, the chapter will bring together literature not often connected— attachment theory and animal communication —and provide a grounded understanding of the infant-directed register in light of both prisms.

As previously noted, attachment theory is the prime scholarly construct when it comes to conceptualizing and investigating human bonding and, as we will see, has in fact largely shaped its contemporary understanding. Investigating what is human attachment and how it works will lead us to the pioneering work of John Bowlby and his development of attachment theory. Equally importantly, this chapter will help us understand how the correct unfolding of attachment can allow for the child's initial intrinsic motivation for proximity to the caregiver (and ultimately for survival) to be progressively transformed into motivation for the exploration of new, surrounding objects. In this vein, concepts such as mentalisation and epistemic trust will prove useful.

As imprinting and attachment make use of species-specific animal signals (called 'releasers') that prompt particular forms of infant engagement, in the second half of this chapter I will also scrutinize them from a communicative point of view. Accordingly, infant crying and some of their ‘cute’ or *neotenic* features will be analysed by means of the concepts presented in the first chapter. The same will be done with infant-directed vocalization, with a natural emphasis on its most studied manifestation: IDSp. Acoustic signals such as infant cry and infant-directed vocalization will lead us to the conclusion that, due to the myriad of selective pressures that bipedalism entailed, human imprinting and attachment became more intensely based on our communicative resources, particularly those of the acoustic domain. Furthermore, these communicative resources will in turn finally lead back to the starting point in these introductory chapters: music.

1. Attachment theory

1.1. Initial definition and theoretical background

Ethologist Konrad Lorenz’s (1970) developed the concept of imprinting, an important process whereby a wide range of animals develop the ability to recognize and become physically attached to companions. The idea of imprinting rests on kin-specific communication such as cooperative breeding and mother tongues, both reviewed in the previous chapter. By the time Konrad Lorenz and Dutch biologist Nikolaas Tinbergen began developing the concept of imprinting in birds, Sigmund Freud had already stressed the psychological impact of early interaction between a mother or father and their infant on the latter’s adult life, including his or her sexual behaviour (see Vicedo, 2009). Thus, a number of psychologists, animal psychologists, psychiatrists and child analysts from different psychoanalytical schools were engaged in studying the way in which a disturbance in an infant’s bond to its mother would imply consequences for his or her adult personality (Ibid.). Bowlby had accumulated a corpus of observational data mainly along with his associate, the social worker and psychoanalyst James Robertson. However, their findings were not yet conclusive, largely since they lacked both experimental examination and a comprehensive theoretical framework (Van der Horst, 2011). In such a context, imprinting theory provided a biological approach to the nature and function of the bond between a child and its caregiver, elements Bowlby needed for completing his formulation of attachment theory, through which he aimed to synthesize ethology and psychoanalysis into a single explanatory framework (Vicedo, 2009).

In the first volume of his classic trilogy, Bowlby (1982) broadly defined attachment as a class of social behaviour (such as mating or parenting) that leads the young human child to maintain proximity to his mother-figure. As such, it is held to have a biological function specific to itself: protection from predators (Bowlby, 1958). Attachment behaviour was characterised by two main features. The first is maintaining proximity to another animal and restoring it when impaired; the second concerns the specificity of the other animal.

In concordance with the development of contemporary ethology, Bowlby considered the innate/acquired antithesis as unsuitable. Instead, he postulated the child's tie to his mother as a product of the activity of a number of behavioural systems that have proximity to mother as a predictable outcome. Bowlby assembled his proposal by drawing simultaneously on analytical biology and control theory which, together, provided basic principles that worked as an alternative to instinct: adaptive, goal-directed behaviour.

The main element that would allow an organism to fulfil its goals is the notion of feedback, understood as a

> ‘process whereby the actual effects of performance are continuously reported back to a central regulating apparatus where they are compared with whatever initial instruction the machine was given; the machine's further action is then determined by the results of this comparison and the effects of its performance are thus brought ever closer to the initial instruction’ (Bowlby, 1982, p. 41).

Although initially conceived within the conceptual world of early cognitive theories, predicated on information theory and cybernetics, Bowlby deemed that there was nothing intrinsically inexplicable when applying feedback and related principles to living organisms. He started from the position that any given biological structure is unintelligible unless it is considered in terms of survival within a very particular environment. Thus, instead of being explicitly designed to operate in a particular environment like a machine is, a given species would gradually become evolved in its environment. This environment within which a given species has evolved was termed ‘environment of evolutionary adaptedness’. Control systems can be linked and integrated, their settings and goals being partially or completely derived from each other. This applies to the goals and environments of the heart, the circulatory system, as well as to the human being they belong to.

The functioning of some systems is more sensitive to changes in the environment than others. Thus, any biological character that in its development is little influenced by variations of environment can be considered 'environmentally stable' (Hinde, 1963). In the contrary case, it can be thought of as ‘environmentally labile’. Such an approach to unlearned behaviour stands as an alternative to the more controversial idea of instinct, and applies to FAPs such as the following response, clinging or smiling. For instance, the fact that a newborn gosling will always display a following response towards the first animated object it sees, implies that such a response— and the system that controls it —can be considered 'environmentally stable', thus depicting a behaviour that will follow some recognisable pattern and that in a majority of cases will lead to some predictable result of benefit to individuals or species. In addition, the sensitivity of a particular system to its environment might also be greater at one phase than at another. As a result, sometimes a given behavioural system is highly sensitive at one phase and then ceases to be so. For these reasons, Bowlby preferred the term ‘sensitive period’ over Lorenz’s original ‘critical period’.

1.2. Criticisms and development of attachment theory

As previously mentioned, in its initial formulation attachment behaviour concerned the alignment of two main elements: maintaining proximity to another animal and restoring it when it has been impaired, as well as the specificity of the other animal. In other words, attachment behaviour implies the development and deployment by the infant of a physical and cognitive capacity for actively and selectively seeking proximity. However, following Bowlby's definition, such a capacity cannot be properly labelled as attachment behaviour—or at least not the kind he focused on and the development of which he described throughout the life cycle—until it is preferentially directed towards a particular figure.

Regarding the first element (a physical and cognitive capacity for actively seeking proximity), Bowlby observes that from the moment it is born, far from being a *tabula rasa*, a human being is equipped with a number of behavioural systems. First, there is from birth a perceptual equipment that tends to orient the infant towards its caregivers. There is also its effector equipment— mainly hands, feet, head and mouth —, also capable of contributing to interpersonal contact. Thirdly, the infant is born with what Bowlby considers as 'signalling equipment': neonatal crying, voicing, and limb gestures[4]. The role of signals— from a communicative point of view —in the development attachment is key to the main thesis and will be scrutinized in detail in this chapter's second part.

On the one hand Bowlby proposed that amongst the behavioural systems mentioned above, there are already some that provide the building-bricks for the later development of attachment. At the same time, however, he already considers them part of a first phase of the attachment process, characterized by orientation and signals with limited discrimination of a

[4] The 'signalling equipment' mentioned by Bowlby consists solely of behaviours (neonatal crying, voicing, and limb gestures). However, as indicated at the beginning of the second chapter, signals can also consist of phenotypic features. Such is the case of pedomorphic ('cute') traits, that will be addressed in section 2.2.2 of this chapter dedicated to the infant-directed register.

figure. During this phase, which lasts from birth to between eight and twelve weeks of age, the infant's ability to discriminate one person from another is limited to olfactory and auditory stimuli. The ways in which a baby behaves towards a person include orientation towards that person, tracking its eye movements, grasping and reaching, smiling and babbling. Each of these sorts of infantile behaviour, by influencing the carer's conduct, is likely to increase the time the baby is in proximity to that companion.

Theoretically, the infant would enter a second attachment phase once orientation and signalling start showing a progressive and stable narrowing down towards one (usually the mother-figure) or more discriminated figures. Bowlby points out that differential responsiveness to auditory stimuli is observable before visual ones. This second phase was theorized to last until about six months of age in normal circumstances. Around that time, two main changes imply the beginning of a third, distinct phase. The first is that maintenance of proximity conspicuously narrows down to one discriminated figure. The second is that proximity-seeking behaviour ceases to be limited to signals due to the progressive mastering of a new behaviour: locomotion. The infant's proximo-distal and cephalo-caudal development extends to include following a departing mother, greeting her on her return, and using her as a base from which to explore. Regarding causation of the narrowing down of proximity-seeking behaviour, Bowlby acknowledges two main reasons. On the one hand, a consolidation in the learning of the individual— as opposed to supra-individual — characteristics of an attachment figure, that directs responses mainly or entirely towards it. A second reason offered is that as mammals (and birds) grow older, the reaction to any strange figure increasingly tends to fear and withdrawal. The relationship between these two developmental processes was not discussed by Bowlby, and he did not take an explicit position regarding their relative importance or mutual causation. In any case, as a result, the welcoming and rather undiscriminating responses to everyone other than the mother-figure decline. Although few individuals will be selected to become subsidiary attachment-figures—preferably those whom the child has got to know whilst in the company of its primary caregiver —, most will not. Thus, strangers become treated with increasing caution, and sooner or later are likely to evoke alarm and/or withdrawal. In addition, between nine and eighteen months, perceptual recognition of an individual objectives, locomotion, and other systems mediating a child's behaviour to his mother become organized on a goal-corrected basis. By this time the infant's attachment to his or her mother-figure becomes evident for any observer. The development of attachment continues through a fourth phase in which, by observing the mother-figure's behaviour and what influences it, a child progressively comes to infer his or her set goals as well as their means for execution, thus forming a goal-corrected partnership. By the time this phase is achieved, the child has built mental representations of the primary attachment figure referred to as 'internal working model', which I will address in detail further below.

Bowlby stressed the affective dimension of attachment, essential components of which are dyadic regulation processes. Attachment theory and the empirical evidence it relied on defended an innate, universal and evolutionarily driven need for interpersonal relationships as the thrust behind intentional action, and the keystone of mental development. Such an understanding of basic human motivation was at odds with the then ubiquitous Freudian alternative, which focused on deeper individual-centred and individual-gratifying drives (Fonagy, 2010). Still, the two approaches differ in rather fine theoretical disagreements, rather than being fundamentally incompatible. Largely as a result of this compatibility, a dominant shift away from ego psychology and into the prism of relationality and intersubjectivity can be observed in contemporary theories about child psychology (Fonagy & Campbell, 2015).

In the case of an infant's attachment to its caregiver, the main— but not exclusive — dynamic will be the former seeking protection and comfort from the latter whenever distress arises. Successful regulation results in what Bowlby described as the infant's feeling of 'perceived safety'. Such a haven of safety is underpinned by neuropeptides previously mentioned (such as oxytocin) that, insofar securely attached individuals are concerned, render affiliative behaviour rewarding and give it an 'addictive' nature (Luyten et al., 2017). If consistently provided, thus in turn generates a sense of basic trust in the caregiver, powered by a mesocorticolimbic dopaminergic reward circuit, as well as hypothalamic-midbrain-limbic-paralimbic-cortical circuit (Ibid.). Unsuccessful regulation, on the other hand, leads to what Bowlby described as separation anxiety, less activation in the caregiver's brain regions associated with the reward system, and insular activation— and area previously linked to feelings of unfairness, pain, and disgust (Montague & Lohrenz, 2007).

The neural circuits putatively responsible for social bonding (such as attachment) are comprehensively linked to those of general emotional response (Coan, 2008; Coan, 2010; Insel & Fernald, 2004). Pair bonding and social affiliation are linked to dopaminergic projections throughout the nucleus accumbens, pre-frontal cortex, ventral palladium, and ventral tegmentum, regions otherwise implicated in responses to rewards and punishments, emotion regulation, motivation, and personality (Coan, 2008; Panksepp, 2011).

The fit between the caregiving behaviour habits of the main attachment figure and the infant's dispositions eventually sediment as relatively stable trait-like individual differences in the latter, usually referred to as attachment styles. Mary Ainsworth experimentally identified three main attachment styles through a procedure known as the Strange Situation (Ainsworth & Wittig, 1969). In it, the child is placed in a room and observed playing for 21 minutes while strangers and caregivers enter and leave, thus recreating the flow of (un)familiar presences normally experienced by children. Attachment styles put forward were ‘secure’, ‘insecure avoidant’, and ‘insecure ambivalent/resistant’. A ‘disorganized’ attachment style was later identified as a fourth alternative (Main & Solomon, 1990), these attachment styles having been widely researched to date (see Pearce, 2009; Rholes & Simpson, 2004). The Strange Situation’s success as a replicable protocol led to its prevalence as a research instrument, but has been criticised for lacking the clinical subtlety of Ainsworth (Bretherton, 2003) and Bowlby’s work (Fonagy & Campbel, 2015). A widespread product of research based on the Strange situation is Mary Main’s Adult Attachment Interview (AAI), a semi-structured interview aiming to infer an individual’s present-day state of mind concerning early experiences with caregivers (Hesse, 1999).

Although describing them in detail will not be necessary here, it remains relevant to mention that attachment styles and their consequences continue beyond infancy. The proportion of the life-cycle during which attachment behaviour is manifested varies greatly from species to species. Bowlby noted that as a rule, it continues until puberty though not necessarily until full sexual maturity. Thus, any form of juvenile behaviour that results in proximity can be considered as a component of attachment behaviour. In the case of human beings, the quality and particularities of a given individual’s attachment will largely determine their personality and social life across the whole lifespan (Young, Simpson, Griskevicius, Huelsnitz, & Fleck, 2017). Early experiences such as attachment styles continue to shape mate attraction, more general interpersonal preferences, and attachment dynamics directed at different kinds of agents through adulthood, and the common neural basis of these phenomena has also been described (Coan, 2010; Gillath, 2015). For instance, dopamine and neuropeptides seem to underpin a number of affiliative behaviours such as caregiving, pair-bonding and sexual behaviour (Neumann, 2008; Insel & Young, 2001).

Like in the case of an infant’s attachment to its caregiver, while regulation was originally characterized as a largely unidirectional dynamic (the primary caregiver), attachment and dyadic regulation processes tend to be more complex in adult relationships (Overall, Fletcher, Simpson, & Fillo, 2015). Typically in early adolescence, progressive insertion in society and the search for identity translate into attachment-related functions being transferred from parents to peers (Hazan & Zeifman, 1994). Adolescents thus first begin to preferentially seek proximity of their peers, then start progressively using them as safe havens, and may finally use them as secure bases for exploration. Something similar but at faster rates happens later in young adulthood, as individuals leave their homes for higher education or other similar projects (Lopez & Gormley, 2002). During most of adulthood, large-scale surveys have confirmed romantic partners as preeminent attachment figures (Doherty & Feeney, 2004). Nevertheless, parents, siblings, children and friends have also been demonstrated to constitute full-blown and even primary attachment figures (Ibid.). During the whole adult lifespan, attachment influences parenting style (Jones, Cassidy, & Shaver, 2015) and partially explains adult psychopathology (Ein-Dor & Doron, 2015).

Bowlby points out that continuity between infant and adult attachment can be evidenced by the circumstances that lead an adult's attachment behaviour to become more readily elicited. For instance, in sickness and calamity adults most likely become demanding of others. Similarly, in conditions of abrupt danger or disaster a person will almost certainly seek proximity to another known and trusted person. Finally, in old age, attachment behaviour may be directed towards members of a younger generation due to the unavailability of members of an older or equal one (for a review see Karantzas & Simpson, 2015). The idea that attachment becomes relevant after childhood, at different points of adult life and again in a person's final years is relevant, as it will help us understand the nature of attachment vocalizations throughout the lifespan.

More recent research has posited a number of limitations in the evidence linking early childrearing and later outcomes, as well as the limited power of prediction that early relationships can at times actually offer (Luyten, 2015). Further refinement regarding the mechanisms of attachment and the tracing of its impact in later life includes genetic influences across the lifespan (e.g. Fearon et al., 2014). The relative influence of early and later attachment figures and styles has been subject of debate. Bowlby's initial focus of attachment on a) a primary caregiver and b) a life-long, determinant effect of the quality of the bond established with this figure has led to a 'prototypical' perspective, later contrasted with an alternative, 'revisionist' approach that highlights the potential transforming influence of later experiences in life and further attachment figures (Fraley, 2002). Attachment 'prototypes' are mainly constituted of non-linguistic representations, procedural rules of information processing, and behavioural strategies (Ibid.) One important argument for early prototypes that remains relatively unchanged is precisely that the primitive quality of their constituting elements (procedural, non-linguistic forms of representation) are harder to access and amend once more sophisticated forms of cognition emerge and partially replace them (Sroufe et al., 1990). Early representations are thus considered to remain relatively fixed after infancy because the child's primary mode of thought has changed, thus perhaps remaining latent but not disappearing (Roisman et al., 2005; Sroufe et al., 2006; Waters et al., 2000).

Revisionist approaches, on the other hand, stress early working models' relative flexibility and potential to be modified when the individual's experiences diverge from existing expectation (Kagan, 1996; Lewis, 1998; Lewis et al., 2000). The revisionist perspective does not necessarily predict stability between infant and adult attachment patterns, because the caregiving environment may change substantially, in a positive or negative manner (Fraley et al., 2011). Individuals can exert a degree of influence over their caregiving environments, selecting those that are consistent with their current beliefs and expectations (Collins, 1996). By this token, stability can be interpreted not as a direct product of underlying patterns, but rather as chains of conscious, here-and-now decisions.

Both revisionist and prototype perspectives consider a place for stability and change in attachment patterns. A meta-analysis integrating 27 effect sizes from 23 studies (total n = 1415) reporting test-retest applications (separated up to 19 years apart) of Ainsworth's Strange Situation concluded that prototypes degree of pervasiveness remains significant, if flexible (Fraley, 2002). Such a conclusion was ratified by a later double longitudinal study (n = 591) on adult attachment, which found prototype-related attachment traits to be even more explicative than other factors as measured by the Big Five questionnaire (Fraley et al., 2011). A further, more comprehensive meta-analysis integrating 127 studies (total n = 21,072) reporting test-retest applications separated up to 29 years apart provides further nuances regarding time windows (Pinquart et al., 2013). In accordance with Fraley's (2002) meta-analyses, data returned moderate average levels of attachment stability, with an overall coefficient of r = .39 between repeated measures. However, measures of stability lost significance when time intervals exceeded 15 years. Similarly, coefficients corresponding to time intervals of less than two years were significantly higher than the equivalent in time spans of five years or more. Coefficients were also higher if attachment beyond infancy had been assessed using representational measures rather than behavioural ones.

The relative importance and stability of early caregiving and the prototypes that the main attachment figures (mainly parents and especially the mother) may generate seem thus undeniable but not absolute. Beyond the exact contribution in later life, the main thesis that will be presented in this dissertation and the studies that will follow it will concern children in their second years of life, ages at which primary caregivers are still paramount, and little space for 'revisions' has taken place so far. For these reasons, and without denying potential changes in working models and attachment figures, in this work I will focus on attachment to the main caregiving figures, especially the mother.

1.3. Motivation for enculturation: mentalization and epistemic trust

As discussed in section 1.1 in this chapter, Bowlby initially favoured an evolutionary and ethological perspective. Such a perspective was at odds with Freudian approaches, which contemplated comparatively-more obscure intra-psychic dynamics that underpin human relationships (Fonagy & Campbel, 2015). Although intra-psychic elements cannot be directly observed and assessed, they allow for an exploration of the human mind less compatible with the rather mechanistic quality of imprinting theory and behavioural feedbacks (Ibid). In any case, Bowlby had not completely dismissed intra-psychic aspects of attachment. In the second volume of this trilogy ('Separation'), he developed the concept of an 'internal working model' (Bowlby, 1980), according to which the child's mental representations of the primary caregiver guide subsequent emotional and social behaviour. The importance of mental states was consistently and progressively defended, and Bowlby and further authors progressively considered levels of mental representation (Main et al., 1985). In fact, all of the research on the stability of attachment prototypes discussed in the previous section rely on mental representations as explanatory tools.

In the domain of mental representations and their role in attachment, a concept that will prove useful in the assessment of this thesis is that of parental reflective functioning (PRF), or parental mentalizing. Mentalizing refers to the caregiver's capacity to acknowledge their child as motivated by internal mental states (i.e. feelings and desires) as well as their capacity to reflect upon their own inner experiences (Luyten et al., 2017), to '*look at oneself from the outside and at others from the inside*' (Luyten et al., 2012, p. 4). It was the observation if child-caregiver interactions through the lens of attachment that inspired the inception of mentalizing: in secure attachment relationships, primary figures are interested in the child's mind, are responsive to their gestures and make efforts to interpret them as meaningful leads to the child's experience (Fonagy et al., 1991). Mentalizing does not only concern caregivers. Caregiver mentalization fosters in turn the development of the child's own incipient faculty for mentalizing (Slade et al., 2005), which in turn is thought to foster their emotional self-regulation, and their general capacity for developing secure attachment relationships.

Because oxytocin also fosters explorative behaviour, mentalizing is also linked to feelings of autonomy and agency (Luyten & Blatt, 2013). This idea was already present in Bowlby' (1982) and Ainsworth's (1991) thinking, who took for granted that attachment security would enhance curiosity and encourage the safe exploration of new, unusual elements in the environment. In their view, a supportive attachment figure would render tolerable the uncertainty and provisional confusion caused by new information. Accordingly, direct links have been found between attachment style one the one hand and novelty seeking, trait curiosity, and exploratory interest in the other (Mikulincer & Shaver, 2007). Such a link between attachment and a progressive interest in engaging with new and unknown elements of the environment will be useful further ahead.

As the child grows older and broadens its contact with society, further agents (e.g. peers, teachers, mentors) will exert an influence on the development of its mentalization (Luyten et al., 2017). However, as discussed in the last section concerning attachment figures, the present work will focus on small children and their interaction with their parents, and so will our focus on literature. As an attribution of mental states to oneself and to others, mentalizing is a construct that largely overlaps with that of a theory of mind (Ensink & Mayes, 2010). Because of its close relationship to attachment and other constructs, and because full-fledged theory of mind does not emerge until around four years of age (Tomasello, 1993), I will focus on mentalizing— and its presence in children around two-years-old and their caregivers.

Mentalizing is thus relevant to our purposes, mainly because it highlights the role of motivation in attachment theory. The construct scrutinizes the degree to which a caregiver acknowledges their child as having an internal, mental life that includes desires, and understands their behaviour as being driven by such. At the same time, the child's degree of mentalizing will give them a sense of autonomy and agency in seeking whatever draws their interest— including music.

Mentalizing added depth to Bowlby's proposal of a relationship between attachment and the emergence in the child of an increasingly autonomous motivation for the exploration of the surrounding world. Further extending this line of thought, the concept of 'epistemic trust'— the capacity to hold others as reliable sources of knowledge —(Fonagy et al., 2015) advanced a more explicit connection between the motivational and the cognitive aspects of such an exploration. Children are often faced to kinds of knowledge that are opaque (not obvious or self-explanatory), or find themselves in need for epistemic vigilance— the caution and discrimination that children deploy during observational learning in order to avoid being tricked or misinformed, intentionally or not (Sperber et al., 2010). These obstacles generate a need to distinguish and rely on a trusted source of communication. To this end, human caregivers have developed communicational strategies that involve ostensive cues such as direct gaze and IDSp (Csibra & Gergely, 2009). Details concerning the use of IDSp as an ostensive cue are relevant to our purposes and will be discussed in the next chapter. For now, let us focus on the idea that attachment provides a major— perhaps the most important — source of trust, and therefore of epistemic trust. After all, it is common sense that children (as do adults) learn and internalize knowledge more willingly when experiencing the instructor as caring for them and having their 'mind in mind' (Fonagy & Campbell, 2015).

Literature suggests that attachment styles indeed have a significant impact on the discrimination of agents as reliable sources of knowledge, with secure attachment experiences entailing feelings of acknowledgement and caring, thus being likely to increase epistemic trust (Luyten et al., 2017). While individuals with secure attachment seem to develop a capacity to distinguish credible sources of information, this is less the case of children with anxious or disorganized attachment styles (Corriveau et al., 2009). In this regard, it has been proposed that the evolutionary function of attachment exceeds 'merely' ensuring the child's safety, and that the attachment system was co-opted by evolution into scaffolding the transmission of cultural knowledge (Fonagy et al., 2007). By this token, the major evolutionary advantage of human attachment would be the opportunity given to the infant to develop understanding,

Thus, epistemic trust underlines the epistemic power that is bestowed upon attachment figures. Amid a virtually infinite range of possibilities, caregivers have a privileged place when it comes to selecting what children should focalized their cognitive resources (attention, memory, and learning) on. In other words, all of the deeply engrained motivational reward systems previously mentioned and originally designed for a dyadic context can be wheeled, channelled so that the child trustfully opens to whatever phenomena, objects or knowledge the caregivers exposes them to. Once again, mentalization and epistemic trust are important because, as discussed in the first chapter (section 4), I am interested in how children become motivated to listen to recorded music— an available cultural product.

1.4. Synchrony, attachment and music

As discussed in this chapter's first section, and much like in the case of other mammals, the birth of a human infant triggers a set of species-specific caregiving behaviours aimed at assuring survival, attachment, and several aspects of general growth and development. As much as the specificity (e.g. face gaze, smiling, IDSp and affectionate touch) and amount of caregiving behaviour, the degree of its coordination with the child's state and signals is of paramount importance (Feldman & Eidelman, 2004). The resulting synchronous exchange where partners timely respond to each other's social cues have been characterized (and already mentioned) as protoconversation (Bråten, 1988) and communicative musicality (see chapter 1, section 2.2). Although the main thesis in this dissertation will deliberately not focus on synchrony, its role in early musical behaviours as the ones that will be assessed justifies a brief discussion.

Interpersonal synchronisation represents a form of joint action; social interactions in which two or more individuals co-adapt to each other by coordinating their actions in time and space, resulting in some environmental change (Knoblich et al., 2011). A growing corpus of empirical work has shown that interpersonal synchronisation increases affiliation between adults (Hove & Risen, 2009; Vicaria & Dickens, 2016), conveying the idea that moving with others makes us feel closer to them. Infants take part in such a relationship between synchronisation and affiliation, either when faced with adult strangers (Tunçgenç et al., 2015), or when assessing other people's degree of mutual affiliation (Fawcett & Tunçgenç, 2017). Inherently affiliative, attachment behaviour has also been studied through the prism of synchronisation, with a focus on both vocal (Beebe et al., 1985; Van Puyvelde et al., 2010) and— mostly —bodily (Biro et al., 2017; Guedeney et al., 2011; Lindsey & Caldera, 2015; Reyna & Pickler, 2009; Schoenherr et al., 2019; Yee, 2015) forms of it.

Synchrony also constitutes a link between music and interpersonal bonds. Within the many shapes that joint action can take, rhythmic entrainment— the joint coupling of two people to a common rhythmic structure —is a central one (Clayton et al., 2005). Rhythmic entrainment entails a temporal and an affective component, respectively concerning the physical metrical structures, and the resulting sharing of affective states (Phillips-Silver & Keller, 2012). This co-occurrence of temporal and affective communion between interactive partners serves thus as the base not just for (proto)conversational exchanges, but also for musical ones. Furthermore, if moving with others makes us feel closer to them, and if music entails rhythmic structures more regular and explicit than those contained in speech or other forms of human interaction, then music should generate at least as much affective entrainment as other forms of interaction do. Different areas of the literature converge in agreement with this logic. First, as discussed in the first chapter (section 2.2), it is the infant's innate musicality and the musical quality of protoconversation that enable early bonding. Second, music indeed has been reported to promote affiliation in children (Cirelli et al., 2014). A third element is the use of music as a therapeutic device for promoting attachment between parents and infants (Edwards, 2011) or adults (Pasiali, 2014).

These forms of synchronous interaction and their bonding quality will later on (chapter 5, section 3.5) prove useful when discussing an important extent of the current understanding of children's engagement with music— and understanding that stresses the role of bodily interaction.

2. Attachment from a communicative point of view

2.1. A case for animal signals and biological communication in human bonding

As so far presented in this chapter, imprinting and attachment are animal phenomena and therefore take place in human development and social life. Through imprinting, we learn and come to recognize the features of our own kind. At the same time, we build a preference for and seek proximity to such elements. Once we progressively narrow down such a preference to (a) particular individual(s), a proper interpersonal bond emerges. This socio-affective dimension and its development throughout the lifespan is what attachment mainly stands for.

It has also been acknowledged that imprinting makes use of species-specific animal signals. As previously mentioned, the role of signals— from a communicative point of view —in attachment is key to the main thesis and will be scrutinized in detail in the following sections, in part through the prism of chapter 2. Although the role of signals in the context of attachment constitutes less of a predominant factor than once thought, its relevance continues to be paramount (Wolff & Ijzendoorn, 1997). These species-specific signals that prompt particular forms of infant engagement referred to as fixed action patterns (FAPs). Because releasers evolved to elicit FAPs, and the latter evolved to promptly run to completion as a response, the accomplishment of such an evolved feedback system corresponds to a case of animal communication (Scott-Phillips, 2008). It is also worth mentioning that these signals and therefore imprinting and the early stages of attachment work by means of influence, and not by the exchange of information that language will eventually enable.

Regarding the signals and behaviours involved in human imprinting, releasers so far briefly mentioned are the auditory stimuli that characterise human caregivers' voice— high-pitched voice in particular —, visual features of its face, as well as the tactile and kinaesthetic stimuli proper of human arms and body. The smile was held as a crucial FAP, an index amongst other infant bodily movements and sounds which stand as attempts to form social interaction. The multimodal quality of human imprinting enables the infant from birth and up to about four or six months of age to learn the morphological and communicative characteristics of the species it is been imprinted into.

At the same time, all of the mentioned releasers take part in the primitive phases of attachment, where the infant is oriented at first towards traits and stimuli, and only progressively towards distinctive people. Signals in proto-attachment and attachment are also crucial. Infants signal demanding for proximity and caregiving through infant crying, and some of their 'cute' features. Because none of these signals and feedback systems have been scrutinized in detail, let us now dedicate a section to those which have been better studied: adult-directed infant crying and infant-directed vocalization. As previously mentioned, the case of acoustic signals such as the infant cry and infant-directed vocalization will naturally lead us to the only remaining element in these introductory chapters: music.

2.2. Attachment vocalization feedback as human imprinting and attachment

By emphasizing the differences in imprinting and attachment theories, the previous section suggests that adult-directed infant crying plays essentially similar yet distinct roles in different stages of infant development. In the following sections I will, to a certain extent, confirm Owings and Zeifman's (2004) suggestion that the attachment vocalization feedback system constitutes a crucial component in the process of human attachment. However, by scrutinising the subject in greater detail, I will specify that it is rather during imprinting and proto-attachment— as opposed to Bowlby's wider attachment process —where such acoustic feedback can properly be understood as an animal communication system. As an implication, I will also suggest that the degree of specificity in the form and structure of infant-directed speech (IDSp) and infant-directed singing (IDSi) indicate that they have signal value in the biological and evolutionary senses.

2.2.1. Adult-directed infant vocalization

Perhaps needless to say, infant cry is an inarticulate form of vocalization, and of course not a form of language, as defined in the previous chapter. Nevertheless, the infant cry has a prosody of its own. As discussed, prosody comprises a relevant form of animal communication, making the mentioned vocalization a relevant phenomenon in this work. Through an extensive review, American ethologist Joseph Soltis analysed the possible function of early infant crying as a costly signal (in Zahavian terms, see section 1.3 in the first part of Chapter 2) (Soltis, 2004). His central claims support the idea that adult-directed infant crying is part of an imprinting dynamic, and that the evolved response to such an honest signal of distress is caregiving behaviour. As a mean for structuring this section, let us outline and supplement Solti's claims. First, he recognizes early infant crying as an important means through which infants can maintain contact with the mother because active proximity-seeking behaviours (Schaffer & Emerson, 1964), such as following, are not yet possible. This confirms Falk's (2004a) arguments that rely on the inability of the human infant to perform the proximity seeking behaviour of clinging.

Soltis' second claim is that although acoustically-distinct cry types reflecting specific needs progressively arise with development[5], specialists seem to agree on the notion that adult-directed infant cry's main function is to maintain proximity to the carer (see also Zeifman, 2001). As Bowlby (1982) points out, infant crying more often than not is effectively terminated by stimuli that, in a natural environment, are almost certain to be of human origin, thus maximizing caregiver proximity. Thirdly, Soltis' points out that infant cry stimuli result in brain activity in areas hypothesized to be involved in mammalian parenting behaviour. Further review of the neural correlates of infant crying and parents' responses to it also portray the highly affective nature of this bonding dynamic; parents experience high levels of distress as a response to that of their offspring, both being simultaneously appeased once contact is resumed. Such neural and affective responses vary according to the attachment style the infant develops (Laurent & Ablow, 2012). On the side of the infant, physical separation from the mother evokes separation calls in a variety of mammalian infants, the acoustic structure of the human infant cry being similar to that of the separation calls of nonhuman primate infants (Panksepp, 1995).

Fourth, human mothers and infants exhibit features of adaptation typical of mammalian species that carry their infants, as opposed to species that tend to cache them for

[5] Bowlby (1982) summarized different sorts of crying, acoustically differentiable given the nature of their origin. For instance, crying from hunger starts gradually and becomes rhythmical, whereas crying from pain starts suddenly and is rather arrhythmical.

long periods of time (Blurton Jones, 1972; Zeifman, 2001). In the latter species, infants have independent thermoregulatory mechanisms, and do not vocalize when separated. In contrast, in carrying species such as humans, mothers and infants are in more continuous contact and feedings are more frequent since independent thermoregulation is poorly developed in the infant at birth. In general, mammalian mothers do not respond to isolation calls indiscriminately, but rather a mother's responsiveness depends on whether the vocalizer belongs to her own offspring (Newman, 2003). The fact that human beings are a notable exception to this trend insofar as caregiving is administered less selectively (Chang, 2013) does not prevent parents from recognizing their own. Sound spectrograms show, indeed, that 'cry-prints' are as distinctive as finger-prints and thus facilitate identifying new-born babies. In her response to Soltis' article, Falk reinforces the idea that infant-directed speech and adult-directed infant cry are complementary behaviours that initially evolved in our hominin ancestors in conjunction with the evolution of bipedalism (Falk, 2004).

It is central for the correct formulation of this dissertation's main thesis to mention that as much as infant crying enhances bonding between caregivers and infants, it can also be involved in conflict between them. In this regard, parent-offspring conflict describes the contrasting fitness interests of parent and infant in terms of parental investment (Trivers, 1974). It predicts that an infant should strive to elicit more investment than a caregiver should optimally provide. Because such an excessive demand would prevent them from investing in existing or future siblings, or attending any other matters, caregivers must assess the costs and benefits of abiding the infant's request and will sometimes decide not to. As a form of management in an assessment-management dynamic (see the first part of chapter 2, section 1.1.2), crying is a self-interested effort to maintain or obtain caregiving by regulating the behaviour of others. Unlike in the case of a neonate, as children grow older and become more intelligent, they eventually come to understand the influential power of their own crying. This would normally occur between the third and fourth phases of attachment (from around 18 moths of age, (see Milligan, Astington, & Dack, 2007), when the infant starts understanding the intentions of others. A case for Machiavellian deception (see the first part of chapter 2, section 1.3) can thus be made once a child is able to foresee the caregiver's reaction to their spontaneous crying, and generates a dishonest signal— exaggerated or faked crying —as a measure for obtaining whatever they seek. Parents must therefore rely on their ability to appraise the honesty of a signal, as well as modulate their own neural (in general) and emotional (in particular) responses to it. Accordingly, sex and parenting experience have been posited as factors mediating response modulation to infant vocalizations (Seifritz et al., 2003).

Thus, this section stresses the evolutionarily-unprecedented importance of this acoustic signal in and infant's survival. Infant crying is involved both in parent-offspring conflict and attachment, the latter remaining as its primary function in terms of seeking proximity to the carer and thus securing caregiving.

2.2.2. The infant-directed register (IDReg)

As mentioned in the previous section, the evolved response to the infant's signalling equipment is caregiving behaviour. Falk (2004a) proposed that the persistent infant cry was prompted by the loss of grasping hands and feet formerly involved in clinging to the caregiver's body. As a response to this change, the special acoustic features of infant-directed speech would have initially been selected to engage, and sooth crying infants that were out of physical grasp. In this socio-affective compensatory measure, IDSp's prosodic utterances act as '*disembodied extensions of mothers' cradling arms'* (Falk, 2004b, p. 462).

Caregivers—specially mothers —not only engage in IDSp, but also in infant-directed singing (IDSi). This vocal activity has been shown to generate even longer sustained attention than IDSp does in the early stages of the infant's development while also facilitating language acquisition (Lebedeva & Kuhl, 2010). The main difference between infant-directed and non-infant-directed singing is that the former has a comparatively-slower tempo and is rendered in a more "smiling tone of voice" (as rated by adults) than non-infant-directed singing (Trehub et al., 1997). As previously mentioned, IDSp typically contains expanded, highly repetitive, and distinctive melodic prototypes. Even more so is the case of IDSi, in which nearly identical pitch patterns and tempo on different occasions have been reported (Bergeson & Trehub, 1999). Accordingly, it has been recently proposed that IDSi was selected by pressures stemming from the dynamics of parent-offspring conflict (Mehr & Krasnow, 2017). As discussed in the last section, infants demand parental investment, an important subcomponent of which is attentional investment. In the case of an infant that cannot yet move and a caregiver that cannot immediately approach it, parents will proceed to vocal engagement. Infants should thus have been under selection to discriminate true signals, to attempt to detect faked ones, and to resist them (Dawkins, Krebs, Maynard, & Holliday, 1979). In other words, infants strive to detect when attention is actually being paid to them. Mehr and Krasnow argue that IDSi's design features conform to Krebs' and Dawkins' criteria for effective, ritualized signals, including redundancy, rhythmic repetition, 'bright packaging' and supernormal stimuli. For this reason, IDSi would stand as a more honest signal (see chapter 2, section 1.3) of attentional investment than IDSp could, therefore being selected in cases of distal soothing, especially in cases of proto attachment (see part I, section 4), when infants have not developed autonomous locomotion.

Because of its central role in the present dissertation, I will dedicate the next, brief chapter to characterize IDSp and discuss its relationship to music. For now, let us finish examining the relationship between the IDReg and attachment.

A considerable number of authors have directly or indirectly related IDSp and IDSi to imprinting and attachment. Some of these have already been mentioned. For instance, in the previous section I examined in detail how infant cry releases caregiving behaviour. The latter often consists of tactile behaviours like cradling and stroking, as well as vocal ones such as IDSp and IDSi. In turn, these measures will cease the behaviour that prompted them in the first place. By this token, an important extent of parent-infant proto-attachment and attachment develops through dynamic and interactive signalling loops of care and cry (Swain, Mayes, & Leckman, 2004). Similarly, the high-pitched voice of IDSp also facilitates smiling and other proto-conversational behaviours in the infant. In turn, the infant's proto-conversational feedback elicits IDSp in general, and dynamically affects the height of its pitch (Smith & Trainor, 2008).

As briefly mentioned in the section dedicated to imprinting, other important releasers of caregiving behaviour are pedomorphic traits. Protruding cheeks, a large forehead and large eyes below the horizontal midline of the skull have been reported to be prioritized by the human attention system (Brosch, Sander, & Scherer, 2007), elicit perceptions of cuteness (Glocker et al., 2009). Accordingly, infants' degree of pedomorphia in their facial traits predicts maternal behaviours and attitudes (Langlois, Ritter, Casey, & Sawin, 1995). Crucially, empirical studies (Zebrowitz, Brownlow, & Olson, 1992) have also shown that pedomorphic traits elicit IDSp. In particular, pedomorphic traits seem to release the positive affective tone that characterises de IDReg (Trehub et al., 1997; de L'etoile, 2006). In this respect, Trainor, Austin and Desjardins concluded that IDSp reflects free vocal expression of emotion to infants, as opposed to the more inhibited expression of affect that characterizes most stances of AD speech (Trainor, Austin, & Desjardins, 2000).

A final argument for considering the IDReg as part of human imprinting and attachment is the progressive acknowledgment of 'secondary baby talk' (Ferguson, 1977):

speech that has the acoustic features of IDSp yet is directed towards adults and elderly people. As discussed in the first part of this chapter, imprinting during infancy has an impact in mating later in life. Similarly, Bowlby and later literature on attachment describe the incidence of early attachment in partner choice and in old age. Accordingly, IDSp directed to friends (Bombar & Littig, 1996) and romantic partners (also called 'loverese') has been documented (Bombar & Littig, 1996; Chang & Garcia, 2011). This altered tone of voice is used between roughly two thirds of romantic partners of both genders, across the lifespan (Chang & Garcia, 2011). By representing the emergence of what can be considered an 'adult-directed variant of IDSp', loverese entails a human instance of imprinting having impact on adult mating behaviour. In other words, loverese seems to be an adult consequence of IDSp, the latter being a releaser once delivered by the carer during imprinting.

It has also been reported that caregivers of institutions for the aged use IDSp when addressing elderly people (Caporael, 1981; Caporael & Culbertson, 1986). Content-filtered samples of this form of IDSp was identified as speech to children by naïve listeners, regardless of the actual age of the targets. Caporael argues that, like pedomorphic traits, the appearance of elderly people would be designed to convey helplessness and thus release caregiving behaviour. A major focus of research on IDSp and IDSi has stressed their role in language acquisition, often not stressing enough IDSp's intimate link with caregiving and attachment. The first scholars to fully articulate and emphasized IDSp's role in processes of intimate psychological connection, not only between mothers and infants but also in other relationships, were American psychologists Meredith Bombar and Lawrence Littig. Having acknowledged that the IDReg was directed towards adult, literate humans, the authors argued that expressing and facilitating intimate psychological connection is a central— if not the main —function of the IDReg (Bombar & Littig, 1996).

In this section, it has been shown how IDReg vocalizations are animal signals involved in feedback dynamics of biological communication, evolved to take part in imprinting and attachment. The latter processes are socio-affective, as they are biological. As in the case of any other animal, human imprinting and attachment make use of animal communication dynamics that are common to mammals and birds, the reason why a neonate can effortlessly take part in them. Thus, IDSp and IDSi are signals with particular acoustic features, evolved to respond to, and elicit, the infant's own set of animal signals such as crying, proto-conversation, and pedomorphic traits. Once again, by animal signals— as opposed to cues —, I mean *'any act or structure which alters the behaviour of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved'* (Smith & Harper, 2003, p. 3). Moreover, attachment vocalization feedback can be considered a form of biological communication. Whereas signalling may occasionally fail, biological communication refers exclusively to the successful completion of a signalling act (Scott-Phillips, 2008). Accordingly, an infant being distally soothed or engaged in protoconversation, as well as pedomorphic traits eliciting IDSp and affective disinhibition, would imply the successful completion of biological communication. Finally, the use of the IDReg in adult relationships, as contra intuitive as may seem for some, proves its attachment value.

3. Closing remarks

Let us recapitulate. The first section, on animal and human communication as well and their interrelationship showed that some of the most important cases of biological communication present in human interaction are intimately related to interpersonal bonding. Attachment theory, the prime scholarly construct when it comes to conceptualizing and investigating human bonding lead us to imprinting theory and the subtle differences between these two constructs. Given that, as discussed, imprinting and attachment make use of species-specific animal signals called releasers that prompt particular forms of infant engagement, I scrutinized them from a communicative point of view. Infant crying and some of their ‘cute’ features were analysed by means of the concepts presented in the first chapter. Same was done with infant-directed vocalization.

A chain of causalities has been outlined. First, bipedalism implied a narrowing of the birth canal, which in turn resulted in human infants being born in a state of secondary altriciality with no precedents in avian or mammal evolution. Thus, secondary altriciality acted as a compelling selective pressure for the infant to find new proximity-seeking behaviours within the context of imprinting. Imprinting and proto-attachment— a human being’s first socio-affective, interpersonal dynamics —, once nested in limb bodily action (e. g. following, clinging) in the case of primates, underwent a phylogenetic shift towards face-to-face communication and its multimodal signals, the acoustic domain (infant cry) being chiefly responsible for distal engagement, and thus constituting the privileged facilitator for face-to-face communication and its multimodal signals. Infant crying was thus selected for its efficiency in fostering and maximizing contact opportunities, given that the child’s volition had a significantly smaller power to secure actual physical contact.

Signals in proto-attachment and attachment are also crucial. Infants signal demanding for proximity and caregiving through infant crying, and some of their ‘cute’ features. Researchers in the field have reported a strong link between infant-directed vocalization and music. Such a link is utterly relevant to the present work, since it entails a privileged intersection between the IDReg —a form of animal communication necessary for attachment, but also the seed of musicality— and music. Therefore, in the next chapter I will examine in detail how and why music and musicality are related to IDSp. This will, in turn, allow me to present a main hypothesis that tests such a relation.

CHAPTER 4. THE INFANT-DIRECTED REGISTER

In this chapter I will address the IDReg (in general) and IDSp (in particular) from several, relevant angles. I will review the understanding of the register from its inception, as well as listing and describing its main features and functions. Following previous discussion on epistemic trust (see chapter 3, section 1.3), a section will be devoted to the use of IDSp as an ostensive cue as described by the theory of natural pedagogy. The second half of the chapter will address the relationship between the register and music, starting by briefly reviewing the different ties between prosody and music.

At this point, the chapter will critically address just how exactly the IDReg and music are associated and enquire whether it is the musical aspects of the former that lend it its characteristic high level of attractiveness. This final section of the chapter will deepen into the features and functions of the IDReg that are more briefly described in the first section. I will aim to scrutinize the 'musical' aspects of the IDReg in a level of detail greater than that offered by the contemporary literature: instead of assuming a general, holistic association—however intuitively convincing such an association might be —I will dissect IDSp and IDSi into their constituent features. This analytical exercise will allow us to identify in precise and distinct terms exactly which amid such features can indeed be considered 'musical', and which should not. The outcome of this first section will be that, more than any aspect of the IDReg that can properly be considered as 'musical, it is the register's characteristic 'loving tone' that seems to capitalize infants and children's attention.

1. Basic characterisation of the infant-directed register

Infant-directed speech is a part of the vocal aspect of the Infant-Directed Register (IDReg) (Ferguson, 1977). Essentially, this register comprises the idea that members of almost all known cultures intuitively change the quality of their interaction when addressing infants (Mithen, 2005), constituting a form of 'addressee effect'— the idea that vocal prosody adapts to the addressee (Fernald, 1989). There is a visual and kinetic component, as human infants also prefer infant-directed action or 'motionese' over adult-directed action (Brand & Shallcross, 2008)[6]. The vocal component of the Infant-Directed Register has been conventionally subdivided into infant-directed singing (IDSi) (Trehub, Unyk, & Trainor, 1993) and infant-directed speech (IDSp)— also referred to as 'motherese' (Fernald, 1985), 'babytalk' (Singh, Morgan, & Best, 2002) or 'songese' (Longhi, 2009).

In contrast to non-infant-directed versions of speech, IDSp's features include elevated pitch, wider f0 contours, slow tempo, enhanced articulation of words, enhanced rhythmicity, and positive affective vocal tone (Trehub, 2016). The latter seems to be rendered in a "smiling tone of voice". Considering the role of smile during face-to-face proto-conversation, the fact that human beings can not only see but also hear a smile (Jones et al., 1991) suggests that a crying infant lying away from its mother could 'hear her smile' through IDSp. Its four main functions during the infant's development are modulating arousal and emotion (soothing[7]), engaging and maintaining the infant's attention, fostering the infant's understanding of another's intentions and emotions (mentalising), and facilitating language acquisition (Fernald, 1991). Some of the most salient prosodic adjustments occur in terms of melodic contours (Masataka, 1999): parents typically use a small repertoire of simple, expanded, highly repetitive, and distinctive melodic prototypes in both linguistic and non-linguistic utterances. Because of these features, IDSp has been described as a communicative

[6] Gorillas have been claimed to also present a gestural motherese (Luef & Liebal, 2012).

[7] Human infants have very little means when it comes to emotional self-regulation. A notable exception are whines, which have increased pitch and a more melodic pattern of rises and falls, thus sharing salient acoustic features with cries and IDSp (Sokol et al., 2005) (Sokol, Webster, Thompson, & Stevens, 2005). As such, whines; be thought to provide the infant with an acoustic resource for self-soothing, just as a caregiver would do through IDSp (Katz, 1999).

device whose main acoustic features resemble music rather than language, and in which the melody *is* the message (Fernald, 1991).

2. Natural Pedagogy. The IDReg as ostensive cues

Let us now return for a moment to the relationship between IDSp and learning that was only hinted at in section 1.3 and explain it in further depth. As language allowed for human culture to grow in range and complexity, human infants were accordingly faced with an ever-growing number of human-made, largely arbitrary objects and phenomena (Csibra & Gergely, 2011). Our species thus developed cumulative layers of social complexity that demanded the transmission of accordingly complex social knowledge (Heyes & Frith, 2014). The way to approach these cultural phenomena or objects often may remain opaque to the observing child in terms of their purpose (teleologically opaque), the role of their different components (casually opaque) and their degree of generalizability regarding other similar situations. By this token, children find themselves in two opposing needs: a need for help from somebody else in learning to navigate opaque cultural products, and a need for epistemic vigilance (Sperber et al., 2010).

As a solution to such a conundrum, Gergely and Csibra suggest, humans evolved a cue-driven form of socio-cognitive adaptation the purpose of which is to secure the transfer of knowledge in these challenging circumstances— the theory of natural pedagogy (ToNP) (Csibra & Gergely, 2009). This theory posits that the act of transmitting knowledge from one person to another is rendered explicit by means of non-verbal behavioural signals produced by the 'teacher', referred to as ostensive cues (Csibra & Gergely, 2011; Russell, 1940; Sperber & Wilson, 1986). Ostensive cues include direct eye contact, turn-taking contingent reactivity and— crucially —IDSp. It was already mentioned in the previous section that one of the functions of the IDReg is to engage and maintain the infant's attention (Fernald, 1991);

lending further detail to this notion, ToNP illustrates the way in which IDSp is capable of triggering a learning disposition in the child, favouring the incorporation of new socially- and personally-relevant information. For instance, 6-month-old infants can follow an adult's gaze-shift towards an object, only when the latter has been preceded either by direct eye contact with the infant or by IDSp (Senju & Csibra, 2008). Similarly a person's object-directed gaze-shifts are more consistently attended to by infants than non-object-directed ones (Senju et al., 2008), provided in both cases that shifts be immediately preceded by direct eye contact between the child and the adult. The exact nature of the relationship between gaze and ostensive cues is not yet clear however, as Senju and collaborator's findings have been contested (Gredebäck et al., 2018).

Csibra and Gergely underline the evolutionary aspect of natural pedagogy by claiming that children display a species-specific sensitivity to ostensive cues, to which infants would attend preferentially, and the impact of which would be readily evident in their subsequent behaviour (Csibra & Gergely, 2006). By this token, and although not explicitly delving into the communicative jargon reviewed in our second chapter, ToNP seems to consider ostensive cues (in general) and IDSp (in particular) as animal signals in a strict sense, and natural pedagogy as a case of biological communication. In other words, ToNP defends that the evolved purpose of ostensive cues as signals is to trigger a specific behaviour: attention and learning. At the same time, ToNP defends that children actually assuming a learning disposition as a response to an ostensive cue constitutes the successful completion of a signalling act. One further reference to animal communication is given by the idea put forward by Csibra and Gergely, that ostensive cues serve the purpose of overcoming potential misinformation. Although the concept is not considered in the inception of epistemic vigilance, (Sperber et al., 2010) misinformation (in general) and intentional misinformation (in particular) are closely related to Machiavellian deceit (see chapter 2, section 1.3). As discussed, honest signals are designed to prevent Machiavellian deceit by being costly, and IDSp is indeed a costly signal, that requires effective (and other kinds of) investment.

The concept of natural pedagogy partially inspired, and is therefore intimately related to, epistemic trust. A child embraces new knowledge signalled by a caregiver's ostensive cues *because* epistemic trust has already been developed between them. The opposite is also true: if the adult is not trusted as a reliable source, the child will not welcome new knowledge is spite of the caregiver's deployment of ostensive cues. In this sense, epistemic trust enables the successful unfolding of natural pedagogy. As the possibility of deception between adults is far from absent, epistemic trust continues to be necessary throughout the lifespan. Nevertheless, it will not always be necessarily accompanied by ostensive cues. For instance, in order for a psychotherapy to be successful, the patient must progressively embrace the therapist— and sometimes themselves —as reliable sources of knowledge (Fonagy et al., 2015). In such an adult context, on the other hand, IDSp will not normally be used as an ostensive cue. Still, in children-related contexts such as the ones concerning this dissertation, epistemic trust and ostensive cues will often take place at the same time and overlap in their pertinence.

Thus, in simple words, parents use IDSp to signal to children that they are required to pay attention and learn. Such a use of IDSp is proportional to the opacity of the situation or object. These ideas will be relevant when considering the appeal of the different elements of our thesis: IDSp, IDSi, and recorded music. Considering that the IDReg is naturally appealing— to the point of scaffolding other purposes —, so will be in consequence IDSi. In contrast, recorded music— and many other forms of art —can be considered significantly opaquer in epistemological terms. It can be said to be teleologically opaque, as children need to be taught its purpose (e.g. enjoyment, distraction, socializing, introspection, etc.). It can also be said to be casually opaque, as children need to be taught how to approach it (e.g.

dancing, doing actions, interacting, feeling, etc.). The opacity of recorded music and the role of IDSp in facilitating children's understating of its potential uses in spite of such opacity will be properly discussed when laying out the main thesis.

3. The IDReg and music

3.1 Prosody and music

Before dealing with the particularities of IDSp, it is necessary to acknowledge that, already at a more general level, vocal prosody is already intimately related with music. In fact, prosody is a term imbricated with music from its very inception, as it derives from the Greek *prosoidia*, *or* 'singing-along' (Atkinson, 2008). Greek grammar included speech sounds, and experts in the field were considered simultaneously grammarians and musicians (Ibid). These grammarians already considered *tonus* (ancestor of the words tone, and tonality) as the standard term for vocal inflection and the most important element of prosody.

Beyond its historical interest, the original sense of this etymology has partially endured the scrutiny of modern science. As physical events, music and speech consist of complex auditory signals that share the same acoustic parameters (pitch, loudness, timbre and rhythm), organized around an underlying structure (Heffner & Slevc, 2015). Parallels between the structure of speech and music were first systematically found in terms of the recursivity (see chapter 2, part II section 1.1) found both in speech syntax and tonality (Lerdahl & Jackendoff, 1983). Lerdahl and Jackendoff's generative approach led to the discovery of further commonalities in the temporal domain, as linguistic stress and musical meter also share similarities (see Palmer & Hutchins, 2006). The generative approach left out of focus further important prosodic patterns involving loudness, pitch, and timing, the consideration of which have more recently led to a more encompassing appraisal of the parallels between prosodic and musical structures (Heffner & Slevc, 2015).

Another crucial connection between prosody and music is their capacity to be perceived as expressing emotions. In this regard, an established approach is that the structures of music show formal resemblances to the structures of vocally-expressed emotions, thus triggering similar responses (Juslin & Vastfjall, 2008). In other words, music can contain acoustical patterns that partially mimic those embedded in emotional speech, as different 'channels' that share a common code (Juslin & Laukka, 2003). In line with this view, it has been found, for instance, that musical training enhances the emotional decoding of speech prosody (Thompson et al., 2004). Similarly, localized brain lesions have been reported to generate impairments in the discrimination and perception of both music and vocal prosody (Nicholson et al., 2003; Patel et al., 1998). Arguments that highlight formal resemblances between music and the vocal expression of emotion draw a causal relationship between the two, as our ability to detect emotion in music relies on the evolution of the expression and perception of emotion (Scherer et al., 2001). Such is the case of the super-expressive voice theory, according to which music's appeal and highly emotive quality partially rest in the fact that trained voices and— particularly —musical instruments generate sonic structures that resemble human utterances while at the same time exceeding everyday vocal capacities (Juslin, 2001).

A crucial implication of the parallels between speech prosody and music is that they make it virtually impossible for un-encultured infants to distinguish between them. As discussed in the first chapter, the very distinction between music and language is largely a culturally-imposed one, such imposition requiring time to be learned. Thus, while human adults process speech and music differently— music processing relying more on the right hemisphere of the brain and speech on the left one (Callan et al., 2006) —, evidence also suggests that these neural specifications take place along ontogenetic development (Scott et al., 2007). In line with the idea of progressive perceptual specificity, some evidence has been reported for major overlapping in the neural activity of infants when faced to IDSp and instrumental music (Kotilahti et al., 2010).

Partially based on arguments such as the ones presented in the above paragraph, it has been further suggested that speech (spoken language) is first seized by un- or incipiently-encultured children as a form of vocal performance, attending to the 'musical' features first (Brandt, Gebrian, & Slevc, 2012). The authors defend the idea that without the innate access to sound that I have referred to as musicality (see chapter 1, section 2), accessing language would be impossible;

> These findings suggest that these discrimination abilities may explain how infants solve the bootstrapping problem— i.e. how to connect the sounds to meaning. Put another way, infants use the musical aspects of language (rhythm, timbral contrast, melodic contour) as a scaffolding for the later development of semantic and syntactic aspects of language. Infants are not just listening for affective cues nor are they focused exclusively on meaning: they are listening for how their language is composed (Brandt, Gebrian, & Slevc, 2012, p. 6).

Thus, prosody naturally comprises a pivotal point between music and language. Such a pivotal character will be relevant when laying out the central thesis, in providing grounds for the idea that— given their formal similarities —children might resort to music as a privileged source of vocal affection and affection-related interaction.

3.2. Is the 'musicality' of the IDReg its most appealing feature?

A large number of authors have located the origin of musicality (a capacity for music) in early communicative interaction (see Dissanayake, 2000; Malloch & Trevarthen, 2008; Papousek & Papousek, 1989; Phillips-Silver & Keller, 2012; Trainor, 1996; Van Puyvelde & Franco, 2015), to the point of defending that '*early caregiving is musical*' (Trehub, 2016).

The main element of early communicative interaction that has inspired its association with musicality is the of the already-introduced vocal component of the Infant-Directed Register (Ferguson, 1977) and its subdivisions: infant-directed singing and infant-directed speech. Beyond singing itself obviously being musical, IDSp has been referred to as 'musical speech' by a number of researchers (Fernald, 1989;Trehub et al., 1997) mainly because of the 'sing-song quality' its exaggerated prosody gives it (Trainor, Austin and Desjardins, 2000).

Statements such as those above more often than not imply idea that it is music that better characterizes the most significant features of the IDReg. They may as well suggest that the musical components of the vocalizations of caregivers constitute an important— if not the main —element behind the IDReg's communicative effectiveness. As mentioned many times already, the IDReg and its constant association with music and musicality seem to comprise a privileged natural intersection between communication, attachment, and music. However, this association must not be taken for granted. On the contrary, I should (as I have done before) critically address it. In order to do so, let us analyse the main features of IDSp and IDSi, the ones they share, and the ones that differentiate them. Having done so, the same task will be undertaken but focusing on their functionality.

Both IDSi and IDSp are forms of early vocal communication that share most of their acoustic and functional properties. In terms of acoustics, similarities include elevated pitch, wider contours, slow tempo, modified articulation of words, enhanced rhythmicity and positive affective vocal tone (Trehub, 2016). All of these qualities are conceived in contrast to non-infant-directed versions of speech and singing. The main difference between IDSi and IDSp is their dissimilar degree of formal prescription; songs make use of prescribed pitch intervals and rhythms to a degree that speech usually does not (Trehub, 2016). In linguistics, the terms melody and speech melody are used interchangeably with intonation (Bolinger, 1986); melody in linguistic contexts typically refers to the overall pattern of rising and falling fundamental frequency (f0) contour[8]. Alternatively, in a musical context a melody must consist of perceptually-precise-enough pitch distances traditionally called musical intervals (‘semitones’, ‘major thirds’, etc.) between successive notes. Thus, melody in speech corresponds roughly to melodic contour in music, the latter referring to the directional patterning of successive pitches (e.g., up–down, down–up–down) regardless of the size of pitch change (Singh et al., 2002). In the case of IDSi, nearly identical pitch patterns and tempo on different occasions have been reported, contrasting with maternal utterances which are comparatively-more variable (Bergeson & Trehub, 2007). The strong formal prescription of a melody has mnemonic implications; a melody is reliably repeatable and therefore more memorable. A relevant example of this is the neonatal response to melodies reported by psychologist Peter Hepper, who demonstrated an increase in body movements as a response to a familiar piece of music for near-term foetuses (Hepper, 1991). The same was found for short speech sequences (DeCasper, Lecanuet, Busnel, Granier-Deferre, & Maugeais, 1994). Most probably capitalizing on this mnemonic advantage, mothers spontaneously impose their own prosodic constraints while engaging in IDSp, generating pitch patterns that are distinctive and consistent at an individual level in the manner of a ‘signature tune’ (Bergeson & Trehub, 2007). Such a vocal phenomenon is, as previously presented, not an original human innovation, but a resource equally exploited by birds and cetaceans.

Differences between music and speech in eliciting and maintaining the child’s attention have also been reported. For instance, Nakata & Trehub (2004) have contrasted the stereotypy and repetitiveness that characterise maternal singing to the greater variability of speech; while the former would sustain the attention of the infant by means of moderate arousal levels, the latter would rather elicit cycles of high arousal, gaze aversion, and re-

[8] In phonetics, the fundamental frequency (f0) corresponds to the lowest frequency of the periodic waveform produced by the vocal chords. Although not technically the same, for current purposes, f0 can be regarded as more or less homologous with the audible ‘pitch’ of the voice.

engagement. Similarly, the highly formal prescription that characterizes Western music could have implications for the elicitation of turn-taking, as performance of a prescribed melody does not— initially —contemplate interruptions, making IDSi less suitable for the emergence of turns[9]. Differences in terms of affective expression seem likewise to distinguish IDSp from IDSi. Fixed lyrics and tunes enable singing caregivers to focus on the expressive aspects of their performance, whereas in speech prosodic form is largely constrained by the referential message. Similarly, caregivers usually handle a rather small repertoire of songs (Trehub et al., 1997)— as opposed to the potential infinite number of sentences that speech generates —, which would allow further emotional ritualization. Thus, although IDSp does conveys emotion (Papousek & Papousek, 1989), it may be less effective in this regard than IDSi, especially in the case of pre-linguistic infants who can access the form but not the content of verbal messages. In terms of overall seconds of attention paid to one or the other, while Nakata and Trehub found that maternal singing elicited preferential attention over speech, no such a bias was found in later research (Corbeil et al., 2015). Further evidence was found against preferential attention to speech or music, and interpreted instead as support for the notion that IDSi is as effective as IDSp in getting and keeping infants' attention (Costa-Giomi & Ilari, 2014).

Probably because of the substantial number of similarities, no clear functional differences between IDSp and IDSi have been demonstrated yet. Anne Fernald summarized the functional stages of IDSp as engaging and maintaining the infant's attention, modulating its arousal and emotion, increasing its understanding of the intentions and emotions of another, and facilitating language acquisition (Fernald, 1991). Fernald's functional summary

[9] This would not apply for improvised or invented ID songs, but these are comparatively-rarer occurrences (Trehub, Unyk, Kamenetsky, Hill, Trainor, Henderson & Saraza, 1997; Bergeson & Trehub, 2007). It could also be argued that it sensitises infants to the features— such as phrase-final lengthening —that are likely to be operational in communicative turn-taking (see Jusczyk & Krumhansl, 1993).

remains largely uncontested, and a growing corpus of literature has highlighted the bonding function of IDSp. This initially neglected function is no doubt relevant enough in the context of this work to deserve its own section. For now, what remains important is that, despite differences in attentional and emotional aspects, no distinctive function has yet been proposed for IDSi that is not also performed through IDSp.

Having reviewed the similarities and differences between IDSp and IDSi in terms of form and function, let us now come back to the idea that music better characterizes the most significant features of the IDReg and scrutinize it. Pitch contours that can be considered as melodies, along with enhanced repetitiveness and rhythmicity are no doubt elements of IDSp that can be reasonably considered as defining elements of music. I have already discussed how preverbal infants, although not able to use words, can nevertheless display sensitivity to and make use of communicative dynamics common to other primates and mammals. Because prosody is one of these dynamics (see chapter 2, part II, section 1.2) and both affective and attitudinal information can be extracted from it, it can be well said that in infant-directed vocalization the melody carries a message in itself (Fernald, 1991). By this token, humans can extract emotion from a succession of pitches, be it in the context of speech or music. Similarly, although in everyday use of language exact repetition of content will be considered redundant in the majority of contexts[10], such repetition is perfectly natural in music (including IDSi) which, as already noted, exhibits a high degree of redundancy as compared with speech. Interestingly, so it is in the case of IDSp where, as previously discussed (chapter 2, part 2,) parents typically use a small repertoire of simple, expanded, highly repetitive, and distinctive melodic prototypes (Masataka, 1999).

[10] Exact repetition, although redundant from a strictly informational point of view, can nevertheless still signal pragmatic content. For instance, the exact same word or phrase can be used to convey emphasis (e.g. “no, no, no”)

Continuing with the features of the IDReg, different from both music and speech in the adult register is the elevated pitch, comparatively slow tempo, hyper-articulation of words (in fact any words at all) and positive vocal tone, which can hardly be held to be defining characteristics of 'music'. In the adult register, the same tune will be found in high and low pitch depending on the singer, as the same theme may equally be played by a piccolo or a cello during a symphony. A slow tune is not more musical than a fast one, and neither is a sweet and tender piece of music when compared to an angry or dramatic one. Finally, music can equally be sung (poetry is a fundamental component of lieder, folksong and pop) or instrumental (as in the case of sonatas, symphonies and *lieder ohne worte*).

Thus, only some features of the IDReg can be said to represent most forms of music, such as melodic pitch contours, and a clearer sense of rhythmicity and repetitiveness. On the other hand, elevated pitch, comparatively slow tempo, enhanced articulation of words and positive vocal tone, can be said to parallel the structure of some particular instances of music, but not of music as a more comprehensive category. Following this reasoning, if it were the properly 'musical' elements of IDSp (melodies, repetitiveness and rhythmicity) that made it functionally valuable as it is, the use of 'actual' music should then be even more effective when aiming to engage and maintain the infant's attention, modulating its arousal and emotion, or increasing its understanding of the intentions and emotions of another. In other words, why use 'music-like' sounds (IDSp) instead of music itself? Similarly, if music is effective enough when engaging infants, why bothering embedding it in the IDReg?

Indeed, as previously discussed, not just any kind of singing is enough for eliciting— and maintaining —infants' attention. There is something special about IDSi other than its 'musicality'. Infants prefer infant-directed versus non-infant-directed play songs and lullabies, and naïve adult listeners can identify IDSi (as opposed to singing directed at no particular addressee) across cultures (Trehub, Unyk, & Trainor, 1993). These infant-directed songs are perceived and rated by adults as more 'loving' than the non-infant-directed versions. Moreover, there is a significant correlation between infant looking preference and adult ratings of 'loving' tone of voice that strongly suggests it is the latter to which infants are strongly drawn (Trainor, 1996). These results suggest that the positive emotion conveyed by the carer's tone of voice stands out as the main salient element for the infant perceptual resources. The same case has been made for IDSp. After comparative acoustic analysis, Trainor, Austin and Desjardins (2000) concluded that IDSp reflects free vocal expression of emotion to infants, as opposed to the more inhibited expression of affect that characterizes most stances of ADSp. In other words, it is the degree of expressed emotion that makes it 'special'. Such a conclusion was strongly supported by Singh et al. (2002), who experimentally tested infants' listening preference by independently manipulating affect (happy, neutral, or sad) and speech register (ID v/s AD). The authors' premise was that studies testing infants' preferences may have confounded affect and register by unintendedly and systematically contrasting positively emotive IDSp with 'matter-of-fact' ADSp. They found that higher and more variable pitch was neither necessary nor sufficient for determining infants' preferences, a fact that was attributed to a more general preference for speech that conveys contrastive positive affect.

Thus, among the above-mentioned common features of IDSp and IDSi, the positive or 'loving' tone of voice (Bergeson & Trehub, 1999; etc.) seems to be one of the— if not the — main factors that explains why infants prefer them over adult-directed speech and singing; the one capable of capitalizing on the infant's attentional biases and prompting its affiliative behaviour. Considering this, it is sensible to inquire what does this vocal quality consist in. Early communicative interaction is an inherently interpersonal phenomenon that largely takes the form of (but is not reduced to) protoconversation (Bråten, 1988). An important aspect of protoconversation is that it consists of the multimodal integration of the different components of the human 'interaction engine' (Levinson, 2006, see Chapter 2). Yet another equally important aspect is the intimate, socio-affective character of this exchange, as highlighted by attachment theory. As previously discussed, infants (mainly through their pedomorphic traits) almost invariantly project a sense of cuteness and helplessness, and elicit a degree of caretaking behaviour. Furthermore, such a behaviour is accompanied by a corresponding alteration of the caregiver's emotional state (Trehub et al., 1997). In other words, one of the elements that infants release in their caregivers is positive affection. As previously discussed, vocalizations index the signaller's affective state, and the more affect it contains, the costlier and therefore honest a signal is. Accordingly, a caregiver's 'happy' voice signals at least some degree of affective— and therefore parental —investment. Indeed, both facial gestures and body posture can change the shape of the vocal tract and therefore the acoustic features of either speech or song. As a result, the caregiver's feelings and actions have potentially audible consequences. Parameters for perception of affect include f0, first and second formant frequencies (f1 and f2) and loudness, among others (Scherer, 1986). For instance, in the case of positive affect such as happiness, facial expressions can affect the frequency of the second and fourth formant, which are raised for certain vowels while the speaker smiles (Sundberg, 1987; Tartter, 1980; Tartter & Braun, 1994). Thus, the multimodal integration of the acoustic and visual correlates of the smile make it possible for human beings— including infants —not only to see but also to hear a smile[11] (Jones et al., 1991) — an expression of positive affect. A caregiver's ability to transmit a smile acoustically to a crying infant lying some distance away illustrates a concrete mechanism, fuelled by multimodality, by which the voice can signal affective investment.

[11] Whether it is the role of the visual aspect which drives the importance of the acoustic correlate or the other way around (Ohala, 1980) remains rather inconclusive. In any case, perhaps direction of causality is nullified by the necessity of co-occurrence (the geometry of the vocal tract means that it's impossible to widen the oral cavity and not raise the 2nd & 4th formants—and vice versa).

The importance of emotional vocalization and the smile in imprinting and attachment led to inquiries regarding the acoustic/spectral properties of the IDReg and its characteristic 'loving' tone. Although the results of such inquiries seem largely inconclusive[12], as an overall appraisal, human sensitivity to spectral information varies significantly from the womb to the neonatal period (roughly the first month of life), and again during the first six months of life. Humans are born largely deaf to spectral properties of sound other than the f0 (frequencies up to 1 KHz), but progressively extend their sensitivity towards upper tones, which are definitively relevant by six months of age. The rising curve of infants' pitch threshold is not clearly defined, as neither is their sensitivity to upper tones. In short, how neonates and infants precisely extract and represent the spectral information of the 'loving' tone has not yet been precisely described and understood.

---

[12] In-utero recordings indicate that practically all of the acoustic energy that reaches the foetus' hearing apparatus is transported by frequencies below 1 kHz, the intensity of higher frequencies (where most formants are located) being severely attenuated by maternal tissue (Querleu, Renard, & Crépin, 1981). The auditory threshold is still immature in the new-born for stimuli above 1.5 KHz but not for lower-frequency stimuli. This threshold can be further lowered due to residual amniotic fluid and remaining mesenchyme in the middle ear space (Sininger, Abdala, & Cone-Wesson, 1997). Prenatal experience with low-frequency characteristics of maternal voices seems to influence early postnatal (38-60 hours-old) perception of maternal voices, with frequencies below 1 kHz being critical (Spence & DeCasper, 1987). While the understanding of the auditory threshold of neonates has arrived at a certain consensus, the development of the threshold from one month onwards is not as clear. For instance, Keefe et al. (Keefe, Bulen, Arehart, & Burns, 1993) have shown that one-month-old infants have 'lower' ear canal conductance for regions above 1 kHz and higher conductance below 1 kHz than older infants do. At the same time, (Cooper & Aslin, 1994) Cooper and Aslin (1994) have shown that infants of the same age still prefer IDSp that contains frequencies above 400 Hz, suggesting that 'lower-frequency' prosodic information was not sufficient to account for one month-old infants' preference for IDSp. Such contrast reveals mismatching definitions of 'low frequency', as well as a hypothetical importance of pitches roughly between 400 and 1000 Hz. The rising curve of the infants' pitch threshold is not clearly defined, as neither is their sensitivity to upper tones. Two and four months-old show behavioural preference for tone sequences with six harmonics (versus one), and four-month-olds' attention increased as a function of the increasing number of harmonics (from no harmonics up to 18) (Colombo, 1985). However, such findings relied on synthetic pitch instead of real voices as stimuli, significantly reducing their ecological validity. Accordingly, after considering these and other similar studies, Kaplan and collaborators concluded that neither frequency modulation in the f0 nor the presence of the first five harmonics alone can account for the response to IDSp in four-month-olds (Kaplan, Goldstein, Huckeby, Owren, & Cooper, 1995). Indeed, in addition to changes in perceptual thresholds, infants in the first months after birth seem to derive pitch information from natural speech in a different manner than do older infants (Cooper & Aslin (1994), only gradually learning to extract and categorize pitch contours, or doing it though different acoustical representations (Bregman, Patel, & Gentner, 2016).

As inconclusive as the psychoacoustic and spectral understanding of the spectral properties of affective vocalization may remain, the notion that a 'loving' tone of voice is a crucial— if not the main —factor behind infants' preference for the IDReg over AD speech and singing stands firm. Interestingly, a far more consequential conclusion can be made based on the discussion so far presented in this chapter. The existence of IDSi— as opposed to adult-directed or 'normal' singing —implies that music *per se*, despite its use of highly defined melodies as well as repetitiveness and rhythmicity, is no different from speech in that it too has to be harnessed in the IDReg in order to maximize efficiency when addressing infants and prompt sustained engagement, and insofar as caregivers intuitively feel like changing registers when singing to them.

Thus, rather than their 'musicality', it seems like it is their insertion in the multimodal, socio-affective and interpersonal, context of early communicative interaction that lends these two forms of vocalization their engaging properties. IDSp and IDSi are— almost without exception —uttered by a human being and apprehended in the context of real-time, face-to-face interaction, whereas 'music' *per se* is not necessarily so. Indeed, as forms of prosody, IDSp and IDSi can be construed as being indexical of the human being who utters them, in the sense discussed in the first chapter; they provide 'information' about the signaller's identity, including features such as gender, attitude, and affective state (House, 2006). Additionally, IDSp and IDSi are indexical in the sense that they provide clues regarding the actual shared communicative situation (Bertau, 2007). Thus, for instance, as a caregiver sings in the ID register, they index (signals) that it is themself— and not a stranger —singing, that their attitude is tender and loving, and that their bond is an intimate one. A different case is that of music, which can be produced by a myriad of means other than the voice, and certainly outside the ID register. Indeed, instrumental music such as the one a piano or a baby toy can produce cannot straightforwardly index the identity of the human being behind them, as vocal music can[13].

Let us, one more time, clearly convey the main idea of this section. Rather than being music and its defining acoustic features that lends the IDReg its attentional effectiveness, it is the communicative properties of the IDReg as an evolved signal that allows music to be steadily delivered to infants. This should not be surprising if one keeps in mind the evolutionary and ethological status of the IDReg. The vocal quality of IDSp and IDSi work as a signal that has evolved in response to sustained selective pressures, being effective in the context of equally evolved communicative feedback dynamics. In particular, it is the 'loving' tone of voice that characterizes IDSp and IDSi that indexes socio-affective 'information', that makes these forms of vocalization highly appealing for infants.

---

[13] It is worth clarifying that instrumental music can potentially index the performer, if one knows their style well enough. For instance, a person can often identify his or her favourite flutist playing, though admittedly neither as quickly as someone talking, and only in certain types of passage. In speech, the identity of the talker (friend (if so who?) or foe) is supremely important (Hawkins, personal communication, 7-3-2017).

CHAPTER 5. VOCAL-AFFECTIVE WEANING AND MUSIC AS A SOCIO-AFFECTIVE COMPENSATORY MEASURE [OR, THE CHICKEN AND THE EGG, LONG VERSION]

All four previous chapters were necessary in order to navigate smoothly into the present one, which constitutes the core of this work. By now, we not only possess a state-of the-art understanding of the IDReg based on a review of the available literature but we can, additionally, understand it as a biological signal, an attachment device, and a natural relative of music. Such a thorough assessment of the IDReg was motivated, let us remember, for its widely spread association with music.

Our interest in the 'loving tone' and its relation to music will allow to reformulate the research questions laid in chapter 1 into new questions concerning the 'loving tone' and its variation through time. Such a variation, largely subject to the child's development, has so far been only tangentially mentioned or studied, and not received direct attention. I will therefore coin a name for it— 'vocal affective weaning' —and attempt to define it properly. The reason behind this effort is that vocal-affective weaning seems to entail a form of continuity between early musicality and later listening to recorded Western music, while at the same time explaining the motivation to do so. I will then re-appraise the problem originally laid out in chapter 1 and present our main thesis as a means for dealing with the problem in its full complexity. Finally, I will sketch a research programme in order to test the validity of the thesis and present a methodological overview of the three studies that constitute the empirical part of this dissertation.

1. Reformulating the problem

In this section, our original questions will be re-appraised and reformulated in light of the sum of the literature that has led to a focus on the IDReg. Although the appeal of the overwhelming majority of music heard around the globe does not depend on the IDReg, such is the case for infants and children up a to a certain age for whom, to an apparently large extent, music is appealing mainly if sung by a loving voice. In the first case, music is an autonomous domain, independent of any particular register. In the second case, music's appeal is largely conditioned to the presence of the IDReg.

Some rather philosophical questions stem from the distinction between these two conditions: can organized sounds whose engaging effects disappear when outside the IDReg be considered as music? Similarly, can organized sounds whose engaging effects disappear when not indexing the human being behind them be considered music? Given that the present dissertation concerns the psychology of musical development rather than its philosophy, we will not delve into such questions[14]. Another, more concrete angle from which to scrutinize such a distinction is human development. If infants do not initially engage with music that is not embedded in the IDReg, when do they start doing so? When does music the timbre of which is not the loving voice of a caregiver starts to be appealing for a child? When and why do parents stop using IDSp and IDSi? In sum: when do infants start to consistently engage with music that is not harnessed into the IDReg?

As will be made evident in the next section, these new questions replace those originally presented by means of the first chapter, the value of which I explicitly established. Our original, more general questions were: where does a child's motivation for listening to recorded, non-infant-directed Western music come from? After having experienced the luxury of infant-directed speech and infant-directed singing, why would a child settle for a

[14] Should these questions be of further interest, the reader can resort to the references in chapter 1, sections 1 and 2. For now, let us simply note that these questions can be dealt with by considering the extent to which the category 'music' is culturally constructed, highly variable and polythetic (i.e., no single feature or set of features unambiguously characterises all musics). Hence any given cultural manifestation of "music" need have no simple relationship to the "musicality" that is manifested in IDSi (and indeed in IDSp).

paler alternative such as a recorded track? At which moment(s) of development do such concession(s) take(s) place? Can some kind of continuity be observed between children's initial motivation to engage with IDSp or IDSi, and their later motivation for engaging with recorded Western music? Are there any extra-musical phenomena that could help explain such a continuity? In consequence, can some kind of mechanistic, observable chain of events link early musicality and later listening to recorded Western music?

All of these initial questions can— at least partially —be answered by means of the new, more focused questions. In other words, I am postulating that asking when and how do Western children start being motivated to listen to recorded music (a process of enculturation) is, to a degree, the same as asking when and how do they start being motivated to listen to music outside the IDReg.

2. The thesis

In this section I will provide a novel, hypothetical answer to the question that seems to lie at the heart of all other questions so far presented: when and why do infants start consistently engaging with music that is not harnessed into the IDReg? I will do so by bringing together all the relevant notions and arguments that I have systematically presented and referenced throughout the previous chapters, assuming they have been duly acknowledged. It is my proposal that the question can be partially answered by examining the interplay between two systems, each in delicate balance: attachment and parent-offspring conflict. Changes in one system affects the other's balance, as well as the role that both the IDReg and recorded music play in them.

As seen in the previous chapter, attachment is a class of social behaviour that leads the young human child to maintain proximity to his mother-figure. It is fundamental to survival in early human life, as timely proximity and caregiving may make the difference between life and death. As a result, the child’s motivation for participating in attachment feedback systems is from the very beginning intrinsic and total.

As discussed, unlike the case of more precocial species, human infants are born in a state of secondary altriciality, which means they are remarkably helpless and dependent. As a first consequence, in terms of parent-offspring conflict, the combination between the child’s demands and its genuine lack of autonomy when it comes to dealing with such demands translates into greater pressure on the caregiver’s part. In other words, the infant can freely demand investment, and the caregiver is largely forced to oblige, without room for negotiation. Applying this to the development of attachment during infancy means the caregivers must greatly invest in affection (among other forms of investment. e.g. attention) in order to maintain the infant’s feeling of perceived safety. As Bowlby indicates, in terms of attachment, such a feeling of safety can be largely equated to receiving affection— felling loved.

A second consequence of secondary altriciality is that human infants are unable to actively seek proximity to their caregivers, thus relying almost exclusively on their communicative capacities for maintaining proximity. Communicative proficiency is thus particularly crucial in the development of human attachment, and even more in the case of vocal communication, which can be equally proximal or distal at a time when the infant is virtually blind and cannot move around by itself. As seen in chapter 2, prosody is a form of communication older than language. As such, children are able to extract meaning from it before they can do so through words, including signs of affection. Therefore, special acoustic signals emerged in the evolution of human imprinting and attachment: infant cry, and the IDReg. Such signals, allow for attachment to develop smoothly, compensating for the infant’s immaturity.

Adding the consequences of secondary altriciality to those of both attachment and the parent-offspring conflict means that the caregiver, in order to maintain the correct development of attachment, will be forced to invest considerably more than the infant, in general. When it comes to affective investment in particular— which is largely conveyed by means of vocalization —, caregivers' vocalization directed at the infant remains, for a certain amount of time, distinctly more affectionate than most everyday adult-adult vocalizations. The generalized use of IDSp and IDSi— both highly affective forms of vocalization —are largely a reflection of this situation.

However, as the infant becomes a child, the situation changes. The infant masters new, more autonomous proximity-seeking behaviours (such as clinging, crawling or walking) while its pedomorphic traits recede. In terms of parent-offspring conflict, these changes imply important shifts of pressure on the caregiver's end. On the one hand, new forms of investment will be added or demanded to increase, as children will need to eat progressively more, need to be educated progressively more intensely, etc. On the other hand, existing forms of investment may decrease or even disappear. In this regard, caregivers can expect the infant to become more active while at the same time being themselves less influenced by its pedomorphic traits. Because of its so far paramount importance, one of the fronts on which parent-offspring conflict will thus manifest in trying to strike a new balance is the quantity and quality of vocal affection. The child will naturally attempt to maintain the high degree of vocal affection that its former helplessness originally granted with a minimum of effort, while the caregiver's space for frustrating such expectation will only grow as the child matures.

Added to such conflict will be the caregiver's progressive attempt to use its voice for enculturation purposes, beyond mere affection and safety. Parents' expectation to go beyond prosody when communicating to their children and gradually use language in its combinatory, referential, comparatively-less affective quality (see chapter 2, part II, section 1) will also increase as the child's linguistic proficiency rises. Indeed, because general enculturation is taking place, language will progressively become the vehicle of normativity— and therefore, of conflict. Vocalization, once predominantly a vehicle for affection, will be more and more intermittently so. Importantly, in all cases, a decrease in the quantity or quality of vocal affection will largely equate with a decrease in the quantity or quality of its paramount devices: IDSp and IDSi.

As previously mentioned, the sum of the changes described in the above paragraphs concerning the affective dimension of caregiver vocalization have only been tangentially addressed, and not problematized or systematized. In order to more easily hypothesize how such changes can explain children's motivation for listening to recorded music, we find ourselves in need of a concept that would allow for these scattered elements of information to come together and form a coherent picture. I will therefore in the next section introduce such a concept and coin a name for it. Doing so will allow us to more easily integrate and problematize the mentioned changes in the caregiver's use of the IDReg into the larger whole of the thesis.

2.1. Vocal-affective weaning

Weaning is defined as the process of gradually introducing a mammal infant (since only mammals produce milk) to what will be its adult diet and withdrawing the supply of its mother's milk (Whitehead, 1985). Weaning comprises but one among many dynamics of parent-offspring conflict (Trivers, 1974), in which caregivers and infants are expected to disagree over how long the period of a given parental investment should last, over the amount of parental investment that should be given, and over the altruistic and egoistic tendencies of the offspring as these tendencies affect other relatives. Following the same principle, vocal-affective weaning (henceforth ‘VAW’) can be defined as

> the process of gradually introducing a human child to what will be its adult, culturally defined baseline intake of vocally-expressed affection.

Thus, VAW is partially a consequence of parent-offspring conflict, as parents are progressively less pressured to invest as the child matures. VAW is also partially a consequence of general enculturation— enculturation let us remember, is largely concerned with norms. Children must eventually enter normativity, and the main vehicle for their normative instruction— and the conflict that will come with it —will be language. This, in the context of toddlerhood or ‘the terrible twos’, a developmental period widely perceived as a time where control is disputed between toddler and caregiver (Gallacher, 2005). As a consequence, parents’ use of language will be increasingly focused on information and the transmission of normativity, and decreasingly on conveying affection through prosody.

There is existing literature that— in an indirect manner —has to an extent described the unfolding of VAW. Such literature contemplates more or less exact developmental windows, and has approached the voice through different, precise acoustic features. Because this literature will allow us to set up a research programme for the empirical study of VAW, I will address it in detail at the end of this chapter. Before that, and having defined VAW, let us finish presenting the main thesis.

2.2 Listening to recorded music as a compensatory measure

At the same time that VAW takes place, attachment and its feedback systems remain, in parallel, active and in need for balance. As in the beginning of its life, the child’s motivation to receiving affection and attention in order to feel safe and loved remains virtually at its highest. Therefore, as parents progressively start partially reducing the use of IDSp for developmental and cultural reasons, the child can be expected to be highly motivated to strive to maintain such an affective standard.

My main thesis is that recorded music— a culturally-sanctioned source of affect — should constitute a means for filling the vocal-affective gap that VAW could generate. This idea follows the same 'communicating vessels' logic as the one used in Falk's theory, insofar it comprises a socio-affective compensatory measure. Falk (2004) proposed that the IDReg emerged as an aural compensation for the impossibility of cradling infants who could not cling and thus had to be soothed distally. In other words, physical contact and the IDReg constitute analogous resources, with different material characteristics— aural versus kinetic —but the same function. In this case, because affection and safety cannot be fostered through vocalization in the same manner as during the first year or so, new behaviours will need to be added to preserve equilibrium in the different attachment feedback systems.

But why would children recur at all to recorded music as a source of affection? Already at this stage of development, children 'use' music for many different purposes other than seeking affection such as emotional expression, aesthetic enjoyment, entertainment, communication, physical response, enforcement of conformity to social norms, continuity and stability of culture, integration of society, 'self-regulation' and identity (Barrett, 2010; Campbell, 1998). This question can be answered from several angles. First, let us examine the question from the point of view of attachment. As explained in chapter 3 (section 1), the establishment of secure attachment and mentalisation enable children to use attachment figures as a base from which to explore the surrounding world. In other words, the object of the child's attention and interest is gradually shifted from the attachment figure(s) towards further people or objects, and the child's level of independence in these explorations also increases. In this regard, while IDSi and IDSp are appealing to the child from childbirth and are familiar even from before (see chapter 4, section 3), non-infant-directed forms of music are comparatively less-familiar, opaquer (see chapter 4, section 2) and will therefore need further exploration. Once again, the motivation for this exploration will naturally stem from secure bonds.

A second angle is that of interaction and autonomy. Literature testifies that children's exploration of music other than IDSi is assisted by significant others and embedded in interaction. As children develop and become enculturated, music constitutes an— important —addition to the repertoire of media for interacting with others (Ilari, 2005). At the same time, musical play acquires meaning largely *because* it relates to these interactions (Marsh et al., 2016; Small, 1998). By this token, while meaningful interactions foster engagement with music, the opposite is equally true. As children develop and their attachment fosters the autonomous exploration of the surrounding world, children's relationship with becomes less mediated by social interaction and more self-focused or individuated (Forrester, 2009). Nevertheless, the absence of actual physical interaction does not rule out interaction still being present in a more internalized, enactive manner in solitary music listening. In this sense, it has been posited that listening to music triggers social cognition, as listeners tend to attribute human intentions and feelings to the music they listen to (Koelsch, 2010) and the latter can evoke implicit affiliation (Vuoskoski et al., 2016). Similarly, 'passive' Western music listening has been argued to contain 'covert performance'— cognitive and motor residues of action, interaction and entrainment (Cross, 2010).

Yet another manner in which social interaction can be present in solitary listening to music is not through cognitive or motor enactment, but through evaluative conditioning (EC). This concept refers to an emotion being induced by a piece of music because the latter has been consistently paired with another stimuli (see Juslin & Vastfjall, 2008). For example, a particular nursery rhyme may be repeatedly heard during pleasurable interaction between a child and an attachment figure and, as a consequence, the child will experience pleasurable feelings when listening to that particular piece of music in a different context. EC is similar to episodic memory (EM), a process whereby an emotion is induced in a listener because the music evokes a memory of a particular event in the listener's life (Ibid.). Furthermore, some evidence indicates that EM evoked by music tends to involve interpersonal relationships (Baumgartner, 1992). The main difference between these two phenomena is that while EM presupposes conscious recollection of previous events, EC seems to be unconscious (De Houwer et al., 2001). Therefore, EM will become more likely to take place as children's mentalisation capacities settle. Thus, according to the arguments in this paragraph, children recur to recorded music as a source of affection because of its latent interactional quality.

A further answer to the question of why children would recur to recorded music as a source of affection will now be added to those above discussed. My answer to such a question lies on enculturation and the music/language dichotomy that characterizes Western society. As explained in the first chapter, children are born with a series of communicative capacities referred to as communicative musicality by Malloch and Trevarthen (2008) and as the human communicative toolkit by Cross and Woodruff (2009). Such communicative resources are neither 'musical' nor 'linguistic' in the infant's unenculturated mind. Indeed, as discussed in the previous chapter (section 3.1), it has been argued that language is first grasped as a form of music through prosody, only later to be used and understood separately from music (Brandt, Gebrian, & Slevc, 2012). Because children are endowed with the cognitive tools to engage with music long before they can access language, the musical quality of prosody— made as obvious as possible through exaggeration in the case of the IDReg —allows them to participate in human interaction.

The gradual distinction between language and music at the ontogenetic level has been theorized before in terms of evolutionary speculation (Brown, 2000, 2001) and developmental modelling (Van Puyvelde & Franco, 2015). This thesis' approach to the matter is seeking to provide an explanation for such a distinction based on enculturation and pragmatics— what is done (or not done) through music, and what is done (or not) through language. It was previously explained that the child is exposed to music from birth mainly through IDSi and the affectionate setting of early interaction, a setting that stands as an attachment device. Through this experience, the Western child slowly learns that music is used in affiliative situations: situations of interpersonal enjoyment and positive affect that are in the overwhelming majority of cases *conflict-free*.

The pragmatics of language, on the other hand, will evidence significant changes subject to the child's developmental state and the degree of enculturation that such a state affords. From birth and during the first months, the child is exposed to language mainly in the form of IDSp and the affectionate setting of early interaction, a setting that also stands as an attachment device. However, as explained, the use of the IDReg and the level of affection in voice eventually recedes, partially because of changes in parent-offspring conflict, partially because caregivers progressively find themselves in the need to use language for the communication of normativity and the conflict that comes with it. An important nuance to be added at this point is that the present argument does not rule out the possibility that the conventions of songs and nursery rhymes may be used by adults to foster children's understandings of cultural conventions (Williams et al., 2015). Once again, the difference is that in such cases music will not convey real interpersonal conflict.

By this token, the Western child slowly learns that language is used not only for affiliation: language— unlike music —will just as naturally become the medium of conflict. To illustrate in simple words: the child will, in theory, never be corrected, scolded or told off through singing. All such instances of negotiation between the child and the caregiver will be done through talking.

Music enculturation, thus, takes place in the context of the larger, general process of enculturation. As the child is further introduced in its culture, it learns that the affective disinhibition displayed in the caregiver's speech prosody during the child's infancy will no longer be ubiquitous, but will occur in culturally-defined situations and for culturally-defined reasons. At the same time, it learns that the pragmatics of music that had already been presented and incorporated during infancy will be enriched and diversified but still remain fundamentally affiliative. The child thus progressively conceives of music as a medium for seeking conflict-free interaction (e.g. jointly music making, singing or music listening with the caregiver), or for enjoying its affective quality in personal consumption. In doing so, the child's musicality has been employed by society to shape an engagement with recorded music at the heart of which lies positive affectivity.

By this token, the child's conception of what music *is* starts culturally narrowing down into Western specificities. Because the motivation to receive affection and attention as forms of attachment dynamics is significant, the child will in consequence be equally-highly motivated to listen to music beyond the IDReg. In other words, because the child has a constant need for affection, and because Western culture dictates that language should stand as a progressively-smaller source of it, the child will settle for alternatives in order to maintain balance in terms of attachment. One such culturally-sanctioned alternative is recorded music, ubiquitous in our society. In settling, the characteristics of the child's engagement with music is further narrowed down into Western standards: from the interactional (Turino, 2008) setting of live IDReg to the presentational mode that recorded music implies; from the multimodal integration of live singing, to aural predominance or exclusivity. It has been argued that one important reason why children turn to music is because it motivates them to participate in physical, emotional and cognitive terms, and that such participation is personally rewarding (Custodero, 2006). In this regard, what I am hereby proposing is that such a participation will not always be possible, and less participatory— or less *observably* participatory —forms of engaging with music will be sought to as alternatives.

I have thus finally hypothesised a novel, mechanistic chain of events linking early musicality and later listening to recorded Western music, a chain that I have stated as potentially observable— and therefore testable. In our thesis, the theory of early musicality and later listening to recorded Western music have been connected using as a link the variation through time of the caregivers' use of IDReg. Because of its importance in the overall thesis, I have given such a variation a name— vocal-affective weaning —as well as defining it as a concept. Furthermore, I have implied a negative correlation: the less the IDReg is used, the more children should develop a motivation for engaging with music beyond the IDReg. Such a negative correlation should also, in theory, entail causality: motivation for engaging with music beyond the IDReg is posited to at least partially depend on the use of the IDReg.

Such a negative correlation, I hypothesize, would become stronger (more pronounced changes should be found) and thus observable at least twice during the lifespan, when attachment feedback systems require new equilibria. The first moment should be around the second year, when children reach major linguistic milestones. The second moment should take place at the arrival of puberty, when the slow path to a potential replacement of the main attachment figure starts.

Before proceeding into operationalization matters, let us insist one more time on the different manners in which the scope of the present thesis and the empirical studies that will stem from it detach from previous research. First, although I am interested in children listening to music, I will not focus on cognitive/perceptual aspects— as has been the case in the study of musical enculturation, for instance —, but on motivational ones. Second, although the literature stresses the role of interpersonal synchrony and musical play in early engagement with music, I have instead chosen to focus on autonomous listening (one that is not directly motivated or initiated by someone else) as will often be the case later in life. Third, regarding a possible impact of parental use of IDSp on children's listening to recorded music, let it be reminded that VAW is not expected to be the sole phenomenon ruling the use of IDSp. Instead, we will envisage VAW in the broader context of interaction as motivation, and other uses of IDSp— notably, the ToNP. In this sense, this dissertation will, at best, provide insight regarding the relative importance of VAW amid further phenomena.

3. Operationalizing the thesis

As it stands, and however coherent in theoretical terms it may be, my thesis remains *for now* largely speculative, and many of its aspects, vague. In the thesis, VAW is postulated as a phenomenon that partially explains children's motivation for engaging with music beyond the IDReg. In order to start assessing it empirically and refine it, there are two main elements to operationalize. First, I must define exactly *what* do we mean when we refer to children's 'engagement' with music. In this regard, a section will be dedicated to infants and children's attention to music. In it, factors mediating attention to music will be listed and a systematic account of them will be proposed, and the notions of initial and cumulative fixation will prove relevant. The next three sections will deal with the temporal dimension of the thesis: *when* does it take place. Again, because VAW has been postulated as the main structuring element and the link between early musicality and recorded music, it will be VAW's evolution through time that we will focus on.

The developmental aspect of this thesis is critical and comprises a challenge much bigger than the one attentional aspects entail. The difficulty lies in the fact that the more slowly a given change occurs, the more spread the datapoints must be in order to assess such a change. In the worst-case scenario, the prosodic variations that VAW entails take place so slowly and smoothly between infancy and puberty that it becomes virtually impossible to asses it by means other than a decade-long longitudinal study, a project that greatly exceeds the scope of a doctoral dissertation. An alternative, more suitable scenario for our research context would be that the unfolding of VAW, instead of occurring in an imperceptible continuum, would also comprise moments of a more pronounced change rate. According to the thesis' reasoning, if more discrete— or even better, quantal —changes take place in the caregiver's use of IDSp, then observable changes should follow in terms of the child's engagement with recorded music. For these reasons, finding specific developmental windows during which discrete or quantal changes take place in the caregiver's use of IDSp becomes a fundamental heuristic and methodological goal, if what I intend is empirically testing this thesis.

Section 3.2 will thus examine evidence that sheds light on the developmental dimension of children's attention to music, which will provide a first developmental window to focus on. Next, and as promised earlier, a necessary step for testing the thesis will be defining at which moment(s) of development does VAW take place. Let us remember that we have taken into account phenomena that take place between the child's six months of age (the peak of caregivers' use of IDReg) and a still undefined number of years later.

As explained previously in this chapter, VAW takes place partially because of changes in parent-offspring conflict that are biologically- and culturally-led, and partially because caregivers progressively find themselves with the need to use language for the communication of normativity and the conflict that comes with it. I will thus dedicate a section to each of such phenomena. Literature concerning the impact of the child's language acquisition on the parent's prosody will prove particularly useful to our heuristic and methodological goals, as they will provide precise developmental windows at which our first empirical research steps could be potentially directed.

Having operationalised all relevant elements of the thesis, I will proceed to choose a first developmental window to focus on, and then to briefly characterize it in terms of what music-related behaviours and dynamics can we expect to be taking place. Finally, a methodological overview will be offered, discussing suitable methodologies to test and further refine the main thesis.

3.1. The what. Infant attention to the IDReg and other stimuli.

In this section, I will aim to define *what exactly* will we be focusing on for the remainder of the dissertation. The question I held as most crucial in dealing with our problem is: when and why do infants start consistently engaging with music that is not harnessed into the IDReg? A first nuance to this question is that music outside the IDReg is *somewhat* appealing almost from birth. As already mentioned, infants as young as 5 months-old show rhythmic coupling in response to music and other metrically regular sounds that are played by a non-human device (Zentner & Eerola, 2010). However— and this is crucial —, the attention infants pay to musical stimuli other than IDSi is no more sustained than that prompted by any salient stimulus. Initially, infants turn towards a stimulus as an automatic consequence of their sound localization ability (Trehub & Degé, 2016). This few-seconds-long engagement has been referred to as 'initial fixation', and typically lasts up to 20 seconds (Mehr et al., 2016; Nakata & Trehub, 2004). 'Cumulative fixation', on the other hand, entails a sustained focus that can last for minutes (Nakata & Trehub, 2004) and virtually for as long as the stimulus remains. I may thus refine our original question, by asking at which moment of an infant's development does music other than IDSi start eliciting cumulative fixation— as opposed to initial fixation.

Furthermore, although it is true that IDSi elicits more attention than normal singing, there are many more factors to be considered than register and age alone. Familiarity of the source is one of them. As suggested by literature in chapter 3 (section 1.3), children do not—and sometimes *cannot* —open to a new phenomenon or object unless such a discovery is scaffolded by a caregiver. Another factor to be considered is whether the source allows for co-adaptation (see chapter 3, section 1.3) or not. In other word, whether the source of the music can interact with the child. Unfortunately, most research in the field offers different, unsystematic combinations of these factors, rendering a rather scattered panoramic view. Let us now review the few studies that indirectly inform our research question, and then try to better systematize their findings in terms of the mentioned factors.

A first important one is familiarity. Corbeil and collaborators contrasted the effectiveness of different stimuli in delaying the onset of distress in infants between seven and ten months of age (Corbeil, Trehub, & Peretz, 2015). Results showed that recorded IDSi sung by an unfamiliar voice was more effective than recorded IDSp uttered by an unfamiliar voice. In turn, these two stimuli were more effective than recorded ADSp uttered by an unfamiliar voice. De l'Etoile (2006) found that for 9-month-olds, live IDSi sung by a familiar voice was as effective as live book reading or toy playing in sustaining infant attention, but yet far more effective than recorded IDSi uttered by a familiar voice.

Although still partial, perhaps the most informative evidence is given by Mehr (2016), who tested 5-month old infants' selective attention to two novel individuals after one sang a familiar song and the other an unfamiliar one. A first group of infants had previously heard the target song from one of their parents' voice. These babies stared longer at the same target melody when sung by a stranger than at the same person singing an unfamiliar tune. Furthermore, the magnitude of such a preference was predicted by the amount of previous song exposure. On the other hand, no effect was observed in infants who had previously heard the same song from a socially unrelated person or a toy. Mehr's interpretation of the results is that melodies produced live and experienced at home by known social partners carry social meaning for infants. This is perfectly compatible with the previously discussed idea that vocal music, as a form of vocal prosody, indexes identity and relational 'information' (in quotes, following discussion in chapter 2, Part I, section 2).

Although instrumental or recorded music cannot be directly indexical as vocal music can (see Chapter 4, section 3.2), it is important to acknowledge that all forms of music can become familiar to some extent. Music with timbers and rhythms that have no physical connection with caring adults can still be highly meaningful in that it relates to interactions with them (Ilari et al., 2009; Marsh et al., 2016) as 'people-associations' (Trehub et al., 2015). These associations might be established by means of evaluative conditioning or episodic memory, as discussed in section 2.2.

Thus, as already hinted, these few, relevant studies present different combinations of factors to consider when assessing infant's cumulative fixation to music. In an effort to further systematize and to appraise them as a whole, Table 1 summarizes factors mediating infant attention to music. These are source (human/instrumental), familiarity of the source (familiar/unfamiliar), degree of co-adaptation (live/recorded), register (ID/AD), and type of vocalization (song/speech). The table also assigns a number to each condition or combination of factors, roughly presented in order of effectiveness according to what existing literature suggests. For instance, studies seem to converge in that the most engaging combination would be Condition 1: IDSi performed live by a familiar person. Accordingly, the least engaging one would be Condition 20: a recording of unfamiliar music performed by an unknown instrument.

Source	Familiarity	Degree of co-adaptation	Register	Type of Vocalization	Condition
Vocal	Familiar performer	Live	ID	Song	1
				Speech	2
			AD	Song	3
				Speech	4
		Recorded	ID	Song	5
				Speech	6
			AD	Song	7
				Speech	8
	Unfamiliar performer	Live	ID	Song	9
				Speech	10
			AD	Song	11
				Speech	12
		Recorded	ID	Song	13
				Speech	14
			AD	Song	15
				Speech	16
Instrumental	Familiar source	Live	n/a		17
		Recorded			18
	Unfamiliar source	Live			19
		Recorded			20

Table 1. Factors mediating infant attention to music. Combinations of these factors are presented in a tentative relative level of attractiveness under the 'Condition' column.

A similar hierarchy of appeal most probably applies to listeners of all ages, to some extent. Most children, teenagers and adults would pay more attention to live music than a record, all the more so if it sung by someone explicitly for us, and even more if it carries a loving tone. The difference is that for us, Western adults, live music is a rarer occasion, and even more so having someone singing exclusively 'for us'. Accordingly, recorded music stands as a perfectly suitable and engaging option.

3.2. The when, part I. Developmental aspects of infant attention to music

This section constitutes a first step in delimitating a developmental window during which VAW could have an impact on children's engagement with recorded music. Because research devoted to the years between infancy and school entry have largely focused on production skills and the non-musical benefits of music training, little is known about the "natural" musical environment of children in that period (Trehub & Degé, 2016). This section aims to overcome such scarcity of evidence by, once again, bringing together literature from different research programs. Doing so will allow us to delimitate a first, if still too broad, developmental window to focus on.

Mehr's (2016) and other's results partially confirm the idea that sung music (such as lullabies and play songs) largely obtain their appeal from voice's indexical quality. However, informative as they are, developmental aspects are not contemplated in these studies. One relevant aspect to consider is that children's perceptual capacity to extract emotional expression from the purely aural domain of music is not present from birth (for a short review in the field, see Nawrot, 2003). Instead, multimodality seems to be a necessary element, particularly in early stages. For instance, Nawrot devised an intermodal matching experiment in which 5- to 9-month-old infants were exposed to happy and sad music along with a dynamic visual display (Nawrot, 2003). As expected, infants preferred the affectively-concordant happy display. However, the same did not occur when it came to the affectively-concordant sad. Similarly, after listening to musical excerpts of recorded instrumental music contrasting in tempo and mode (major/minor), 3- to 4-year-olds have been reported as unable to distinguish 'happy' from 'sad' on any basis, while children approaching the age of 5 can, based on tempo (Mote, 2011). On the other hand, by between 6 to 8 years of age, most children have developed an adult-like capacity to extract emotion from exclusively aural music (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001).

These studies indicate that visual stimuli can scaffold infant interest in music that is not IDSi by around 3 years of age. This notion is corroborated by Alexandra Lamont's 2008 study on young children's musical worlds, in which she captured up to 21 daily episodes within a seven-day period of 32 children aged 3.2–3.9 years-old both at home and nursery school (Lamont, 2008). Out of a total of 437 episodes, 81% had music exposure, either at the time of the call (38%) or during the two preceding hours (43%). As Lamont reports, the main activity which accompanied the 353 music episodes was 'entertainment', occurring 30% of the time and mainly consisting of watching television or videos. The next most common activities were 'general play' (17.3%), 'maintenance' activities such as eating, preparing for bed or getting dressed (13.7%) and 'music play' (10.2%), defined as music games at nursery, music classes, children's singing games. There was only one episode of music listening without any other activity taking place (0.3%). Interestingly, Lamont reports a scarcity of live music from others performed to children; out of only three live music episodes, one child was sung to (her father sang to her every bedtime); one child stopped to listen to a street busking group, and a third heard his sibling playing the piano. The other live experiences comprised the children themselves engaging in musical production (singing or playing instruments), either alone or more typically with others. Most music heard was recorded (43.3% through audio recordings and 39% through multimedia). Most children had access to the equipment required to play music and had their own tape players or CD players.

Thus, the studies in this section, although do not inform us directly about the precise degree of attention that children pay to music, do provide important insight regarding the developmental dimension of our inquiry. On the one hand, we know that infants during their first six or seven months of life typically pay a few seconds of attention (initial fixation) to music that is not harnessed into the IDReg and accompanied by concordant multimodal stimuli. On the other hand, we know that infants are progressively capable of making sense of music by relying on its multimodal aspects, so that by during the third year, they are being mainly exposed to and responding to music outside the IDReg. If infants are no different from adults in preferring a loving tone of voice over recorded or instrumental music, what motivations and/or circumstances could lead infants to pay attention to other forms of music? Equally importantly, is there any critical period of transition in this respect between 6 months and 3 years of age? The diminishing curve of the use of the IDReg is mainly unknown, and the potentially-assessable developmental windows between 6 months and 3 years of age are numerous.

3.3. The when, part II. The temporal unfolding of VAW

Reviewing the literature concerning infant attention to music has served two purposes. On the one hand, I have precisely defined what we will henceforth understand when we say that children 'pay attention to' or 'engage with' music. On the other hand, the literature has provided a first delimited developmental frame to focus our research on— the one concerning infant attention to music. Having such an initial overall frame, our next goal is to further narrow it down so as to decide at which moment of children's development I should empirically search for the unfolding of VAW.

So, exactly when and how does vocal-affective weaning take place? The numerous studies that have shown that newborns and infants prefer listening to infant- over adult-directed speech samples have focused on demonstrating how early this preference can be found. For this reason, very little evidence regarding how long the preference lasts has been gathered (Soderstrom, 2007). Evidence of the progressive shift of caregivers from full-fledged IDSp to ADSp derives fundamentally from two different research programs: sex and gender differences in parenting, and the impact of the child's language acquisition on the parent's prosody. Sections 4.3.1 and 4.3.2 consider these two fields that constitute the closest available descriptions of the VAW process.

It is worth distinguishing developmental changes in parental use of IDSp from developmental changes in the children's attention or preference for such a vocal register. The current picture of how infants and children's attention to different sources of vocal input (e.g. IDSp vs. ADSp) changes over time is rather limited (Soderstrom, 2007). Substantial evidence was presented for the existence of a U-shaped curve in three steps: from birth, infants tend to prefer IDSp over ADSp; 7- to 9-month-olds infants show no preference for ID speech; and 10–14-month-olds seem to regain the initial preference (Hayashi, Tamekawa, & Kiritani, 2001). However, such a U-shape dynamic has been challenged, as a more recent study did not replicate the third stage recovery (Newman & Hussain, 2006). Thus, although the mentioned studies do suggest that the preference for IDSp varies as children grow older, further evidence is needed in order to determine the exact ages and nature of such changes.

3.3.1. Sex and gender differences in parenting

This section presents an initial area of available evidence that— indirectly — describes VAW. The research contained in it illustrates how, depending on the caregiver's sex and gender, biological and cultural constraints inflict dissimilar amounts of pressure on the caregiver's role in parent-offspring conflict. Let us remember that the use of IDReg constitutes a form of investment, which means that children will naturally demand it. However, the caregiver's sex and gender will have an impact on the degree to which parents continue to comply with such a demand as the child grows older.

As discussed in chapter 4, the sex and gender of a caregiver has an effect on its response to the signals of the infant. Such is the case of the influence of pedomorphic traits in caregiving behaviour. Empirical evidence suggests that women show greater sensitivity to them than men (Sprengelmeyer et al., 2009). Furthermore, such a difference would be enhanced in women taking hormonal contraception and reduced in postmenopausal women, evidencing a role of the progesterone and oestrogen in mediating such a sensitivity. In a subsequent experiment, while women could consistently prefer the 'cuter infant' (as defined by inter-judge ratings), men had more difficulty in doing so. At the same time, when shown the exact same face pairs but asked to choose the happier or the younger infant, no sex differences were observed. Researchers concluded that the sex difference in the ability to discriminate cues to cuteness in infants underlies female-specific emotive responses (Lobmaier, Sprengelmeyer, Wiffen, & Perrett, 2010).

A similar case has been made concerning the vocal domain. The first tangential approach to VAW can be traced to a gender interest in the intonation patterns of child-directed speech (Warren-Leubecker & Bohannon III, 1984). The authors studied 16 mothers and 16 fathers by recording dyadic sessions with their children (8 5-years-olds, 8 2-year-olds; half boys, half girls), and with an adult. They found that mothers raised their pitch (in contrast to adult-directed levels) equally for both ages of child listeners, but increased their ranges more when speaking to younger children. Fathers increased their pitch and ranges even more than mothers, when addressing the younger children but— crucially —did not differentiate between 5-year-old and adult listeners. In the discussion, the fact that fathers in the sample did not continue to use the same intonation patterns with 5-year- old listeners was related to cultural sex-role expectations. Their reasoning was that with increasing age of listener, sex-role expectations may increase in importance. Given that varied intonation patterns are often perceived as 'emotional' (McConnell-Ginet, 1978), Warren-Leubecker and Bohannon concluded that men quickly revert to the more monotonic patterns typical of their adult-addressed speech to avoid such stereotypically feminine speech patterns. Thus, this study provided the first evidence for VAW; the notion that male parents would restrict their intonation when addressing children 5 years-old rather equally than when addressing adults.

More gender differences in IDSp were found by Foulkes and collaborators (Foulkes, Docherty, & Watt, 2005), who studied segmental features of child-directed speech[15] in a corpus drawn from 39 mothers. Effects were found with respect to the age and gender of the children being addressed. Speech to girls generally contained more standard variants than speech to boys, which, by contrast, contained higher rates of vernacular variants. It has previously been claimed that modifications made in the IDSp register help children to learn linguistic structures and also to learn that speech is a social activity. Foulkes' findings add to Warren-Leubecker's gender-based differences in intonation the notion that IDSp also provides children as young as 2 years-old with differential opportunities to learn the social-indexical values of sociolinguistic variables.

Warren-Leubecker & Bohannon's results and conclusions were more recently confirmed by a study that examined childless adolescent's IDSp (Kempe, 2009). The study explored the link of Empathising Quotient (EQ) and child-related attitudes to the prosodic characteristics of IDSp. Pitch, pitch range, and phoneme durations were measured in 44 females and 45 males addressing an adult and a 4-5-months old infant. The results showed that IDSp increases in pitch and pitch range were similar in males and females while phoneme lengthening, particularly fricative lengthening, was higher in females, which may point to a greater propensity to adopt a soothing intonation. In the same vein, in females, higher EQ was associated with an anticipatory rise in pitch suggesting greater sensitivity to the presence of an infant even outside of child-directed interactions. More importantly, Kempe's results suggested that in young men, greater social awareness was linked to avoidance of 'potentially embarrassing' prosodic features. This finding echoes those of Warren-Leubecker & Bohannon's, insofar it illustrates how prosodic differences can be

[15] In this study as in many others, the term 'child-directed speech' (CDS) is used alternatively to the traditional IDSp.

imposed by cultural sex-role expectations. Penton-Voak and collaborators (Penton-Voak et al., 2007) had shown that 18–24 year old childless men whose faces were rated as higher in attractiveness and prosociality reported liking children but did not increase their pitch as much when addressing them. These findings were thought to reflect a trade-off between mating effort and parental effort, as less attractive men invest more in their offspring than more attractive men (Gangestad & Simpson, 2000). Accordingly, Kempe deemed that male negative relationship between EQ and pitch increase in IDSp corroborated the negative relationship between ratings of prosociality and IDSp pitch reported by Penton-Voak and collaborators. Male vocal inhibition would come about since socially more competent males may exhibit greater awareness of the various social connotations of IDSp and may avoid producing high pitch which could be perceived as feminine and socially awkward.

Some of the phenomena reported in this section constitute examples of audience effect (Zuberbühler, 2008). Audience effects refer to triadic situations that involve a signaller, a receiver and an untargeted bystander, where bystanders affect the signalling interactions between the two individuals. Thus, the presence of a colleague or a distant friend can counteract the disinhibition that pedomorphic traits would trigger in dyadic situations and prevent the father of a 5-year-old from engaging in IDSp. In all of these cases, there is an equilibrium between sex (i.e. hormones), gender (cultural beliefs about appropriate behaviour) intensity of attachment (whether the adult is the primary caregiver or not) and context.

3.3.2. Language acquisition

As explained throughout this chapter, a second reason for VAW to take place is that caregivers progressively find themselves with the need to use language— and prosody —for the communication of normativity as well as handling the conflict that comes with it. This section thus presents a second area of available evidence that— indirectly —describes VAW.

As discussed in chapter 2, part 2, section 1, human vocal communication is a multi-layered phenomenon. I characterized language as a referential, combinatorial structure capable of enormous flexibility and complexity, its linguistic meaning residing in the logical relations between propositions and practical reality. At the same time, language is uttered by means of phonation and thus accompanied by prosody. Unlike language, prosody, does work under the rules of animal communication. The research contained in this section illustrates how caregivers' prosody is systematically affected by the linguistic development of their children. More precisely, it will show how both the quantity (how often it is used) and quality (how exaggerated it in terms of acoustic features) of a caregivers' IDSp change as their children reach major linguistic milestones. When the child was an infant, communication and attachment depended solely on signals and biological communication (see Chapter 2, section I), making the caregiver's prosody the figure and language the ground. Thus, in order to maximize communication, prosody is exaggerated as found in IDSp. As the child learns to talk, language becomes progressively the figure, and prosody recedes towards the ground. Accordingly, the caregiver finds themselves with less need to rely on prosody and its exaggeration. In other words, IDSp is gradually replaced by ADSp.

As previously discussed, language acquisition is universally scaffolded onto the IDReg, given the latter's hyper-articulation of phonemic sounds. Accordingly, scholars have studied IDSp as a facilitator of lexical, phonological, and syntactic learning (for reviews, see Hoff & Naigles, 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Soderstrom, 2007). Amid such a corpus, three studies prove to be particularly relevant given the precise developmental windows they report.

First, Farran and collaborators studied IDSp across 19 American and 19 Lebanese mother-infant dyads, with particular focus on the differential use of registers within IDSp as mothers interacted with their infants ages 0–24 months (Farran, Lee, Yoo, & Oller, 2016). After criticizing a 'monolithical' conception of IDSp in the literature, the authors divided traditional IDSp into infant-directed 'baby register' (IDSP/BR), and infant-directed adult register (IDSp/AR), the second one simply consisting of speech that is directed to infants while still having the acoustic properties of ADSp. The potential importance of the distinction is clear: parents across cultures do not speak with their children using the same vocal range or register throughout the day. Results showed considerable usage of IDS/AR (>30% of utterances) and a tendency for Lebanese mothers to use more IDSp than American mothers. Figure 1 shows the average rate in Utterances per Minute (UPM) of maternal IDSp/BR and IDSp/AR, by age. Both groups showed more IDSp/BR in Utterances per Minute than IDSp/AR. The data also show that IDSp/BR was greater at younger than older ages and vice versa for IDSp/AR.

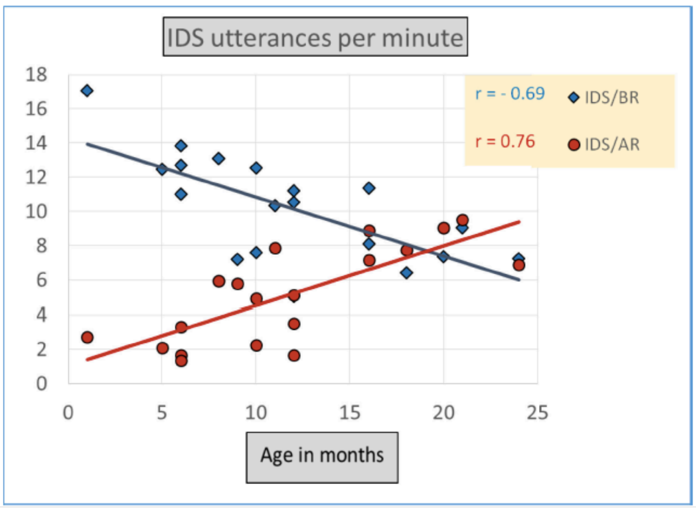


Figure 1. Average rate in Utterances per Minute of maternal IDS/BR and IDSp/AR by Age. Both groups showed more IDSp/BR in Utterances per Minute than IDSp/AR (taken from Farran et al., 2016).

Using UPM as a proxy, the data shown in Figure 1 shows a concrete transition with a fast change rate. The question for cases older than 2 years-old remains unresolved. A crucial change in the BR/AR proportion seems to occur between 15 and 25 months, suggesting that a considerable change in the amount of affect in speech could be expected around that time. It is relevant to note that Farran highlights dichotomic code switching (Bail, Morini, & Newman, 2015) between ID and AD, excluding the gradual change that IDSp itself may evidence.

The next significant work is a longitudinal case-study that scrutinized the IDSp of the mother, father and female nanny of a child from ages 9 to 24 months (Vosoughi & Roy, 2012). Significant interactions were found between speech type and mean vowel duration, f0 mean and f0 range, but not for intensity measurements. Figure 2 shows change in vowel duration, f0 mean and f0 range of all speakers from 9-24 months in child-directed speech and adult-directed speech. The x-axis of the graph is the age of the child, the y-axis is the average f0 mean, the width of the rectangles represents vowel duration and the height of rectangles corresponds to f0 range.

Although a rather a gradual 'adultization' of prosody can be evidenced, a discrete change seems to occur around between the periods 2 and 3 (between 225-430 and 430-525 days-old, or between 0.9-1.1 and 1.1-1.4 years-old). The main variable behind the gap is average f0 mean. This is interesting because, unlike in the case of Farran's study, this gap occurs within the characteristics of IDSp/BR, despite the fact that IDSp/AR can be used as well. The generalisability of these results is constrained by the fact that they derive from one single case-study.

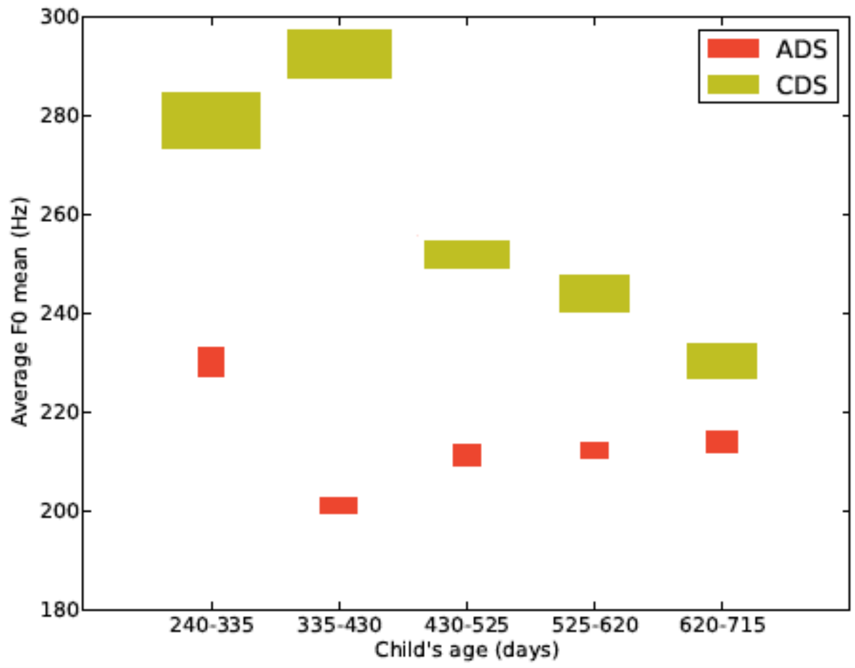


Figure 2. Change in vowel duration, f0 mean and f0 range of all speakers from 9-24 months in CDS and ADS. The x-axis of the graph is the age of the child, the y-axis is the average f0 mean, the width of the rectangles represents vowel duration and the height of rectangles represent f0 range (taken from Vosoughi & Roy, 2012).

Finally, perhaps the most detailed study is the one by Eon-Suk Ko, whose main thesis is that there are quantal changes in speaking rate around the time children reach major linguistic milestones (Ko, 2012). The developmental path of IDSp speaking rate was analysed in 25 mother–child pairs from longitudinal corpora in the CHILDES database. A parallel analysis was also made on the development of speaking rate in the child as well as the mean length of utterance (MLU) in mother and child. The findings revealed that IDSp speaking rate changes nonlinearly with a shift occurring early in the multiword stage. There is also some indication that, as might be expected, another breakpoint might be present around the onset of child speech production.

As seen in Figure 3, In all of the dyads the slope before the breakpoint was positive with p-values at a significant level, meaning that IDSp speaking rate steadily increased along with child age up until the breakpoint. Beyond the latter, however, mothers showed different behaviours, which suggests that the slope depends on several other factors, such as attachment and personality.

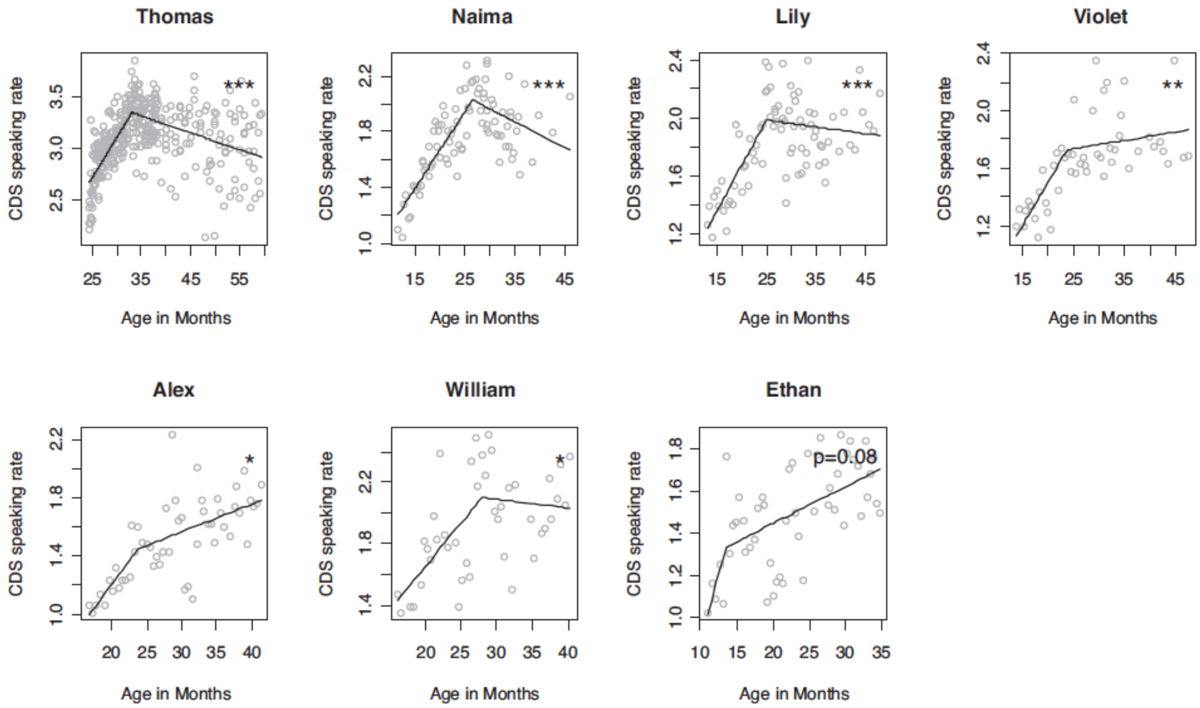


Figure 3. Breakpoint analyses of the IDSp speech rate of mothers as a function of child age (taken from Ko, 2012). Breakpoint analysis of the CDS speaking rate of mothers as a function of child age. The significance codes (*** p < 0.001, ** p < 0.01,* p < 0.05) are from the ANOVAs comparing the breakpoint model with a simple linear regression model.

The methodological approach in this paper is relevant insofar as the focus of investigation includes the shape of the longitudinal changes in IDSp, rather than merely testing correlations. Another important point is the study's attention to individual differences (see Table 2). Mothers and children seem to vary not just in the timing of their developmental changes, but also in the specific implementation of their speech. The average age of the most significant breakpoint reported is 26.2 months, suggesting a change in other parameters around that time.

Child	Age at breakpoint in days (months)	Child MLU at CDS speaking rate breakpoint	Mother MLU at CDS speaking rate breakpoint
Thomas	1007 (33.0)	2.377	5.373
Naima	807 (26.5)	3.917	5.634
Lily	761 (29.3)	2.448	5.324
Violet	895 (23.7)	2.437	5.484
Alex	724 (23.7)	1.32	3.874
William	851 (27.9)	2.442	4.895
Ethan	412 (13.5)	1.1	2.829

Table 2. Child age, child MLU, and mother MLU at the breakpoint in each child (taken from Ko, 2012).

3.4. The when, part III. Defining a critical window for an empirical assessment of vocal-affective weaning

As shown in the previous section, evidence of the progressive shift of caregivers from full-fledged IDSp to ADSp derives fundamentally from two different research programmes: gender differences in parenting; and language acquisition. Both research programmes have reported elements that seem to partially support the VAW hypothesis. Developmental landmarks of the Infant seem to have an impact on the register that is used to address it, as do the culture, sex, and gender of the caregiver. The inhibition displayed by male adolescents and adults is in line with Trainor's proposal of a cultural progressive vocal-affective inhibition (Trainor et al., 2000), and inflect the proposal with gender nuances.

The manner in which changes in IDSp have been assessed are different and complementary. Farran et al. (2016) dichotomize between IDSp/BR and IDSp/AR. On the one hand, by acknowledging that parents across cultures may change register (code-switching) throughout the day, they exemplify a manner in which parents can reduce the prosodic features that convey affect in a dichotomic manner. On the other hand, Farran's proposal does not address the gradual change that IDSp/BR itself may evidence through time, which is precisely Ko (2012) and Vosoughi and Roy's (2012) contribution.

A 'quantal' change is reported by Vosoughi & Roy around the first year of life, primarily relying on f0 data. In Ko's work, on the other hand, the 'quantal' change is reported around the second year of life, by focusing on speech rate. Although both studies' results deserve attention, the fact that Vosoughi & Roy's report is based on a case study makes it less generalizable. It could be the case that the infant they studied had the same outlying developmental curve as Ethan did in Ko's article. Ko's results do match Farran's in signalling the second year of life as a breakpoint, both having a larger sample and using speech rate as proxy. The difference between the two studies is that while Ko reports a shift in IDSp's speech rate, Farran reports a dip in the use of IDSp as a whole. Both results could be compatible.

Hoff and Naigles (2002) stress the idea that both the motivation and occasion for language use that children require lie in social interaction— a context rich in language-advancing data. More specifically, they suggest that between 9 and 18 months of age, the greatest source of variability affecting language development might arguably correspond to that of joint attention. Thus, in this developmental window, even children that have developed comparatively-slower acquire this capacity, smoothing earlier differences in engagement capacities. More importantly, as children level in their competence at staying engaged and following someone else's focus, individual differences in maternal responsivity become progressively-less important. In its place, the amount and informativity of available data is what becomes relevant. This argument can thus be considered another reason for paying greater attention to the turn of the second year of life over the first one, as a universal breakpoint caregiver's use of IDSp.

Thus, available data do converge in defining a critical window in the unfolding of vocal-affective weaning. A transition should be occurring in the use and quality of the IDSp' of parents whose children are roughly between 1.5 and 2.5 years of life, a period during which there is evidence for both an intensification in code-switching between IDSp and ADSp, as well as a significant change in the acoustic features of IDSp itself around that time (see Figure 4).

The thesis then remains that there may be quantal changes in caregivers' use of the IDReg (and its characteristic positive affect) between 18 and 24 months of age, time around which children reach major linguistic, socio-linguistic, social or developmental milestones. The thesis also still holds that there might be an increase in the children's sustained attention to recorded music (music outside the IDReg) as a means for compensating these quantal changes.

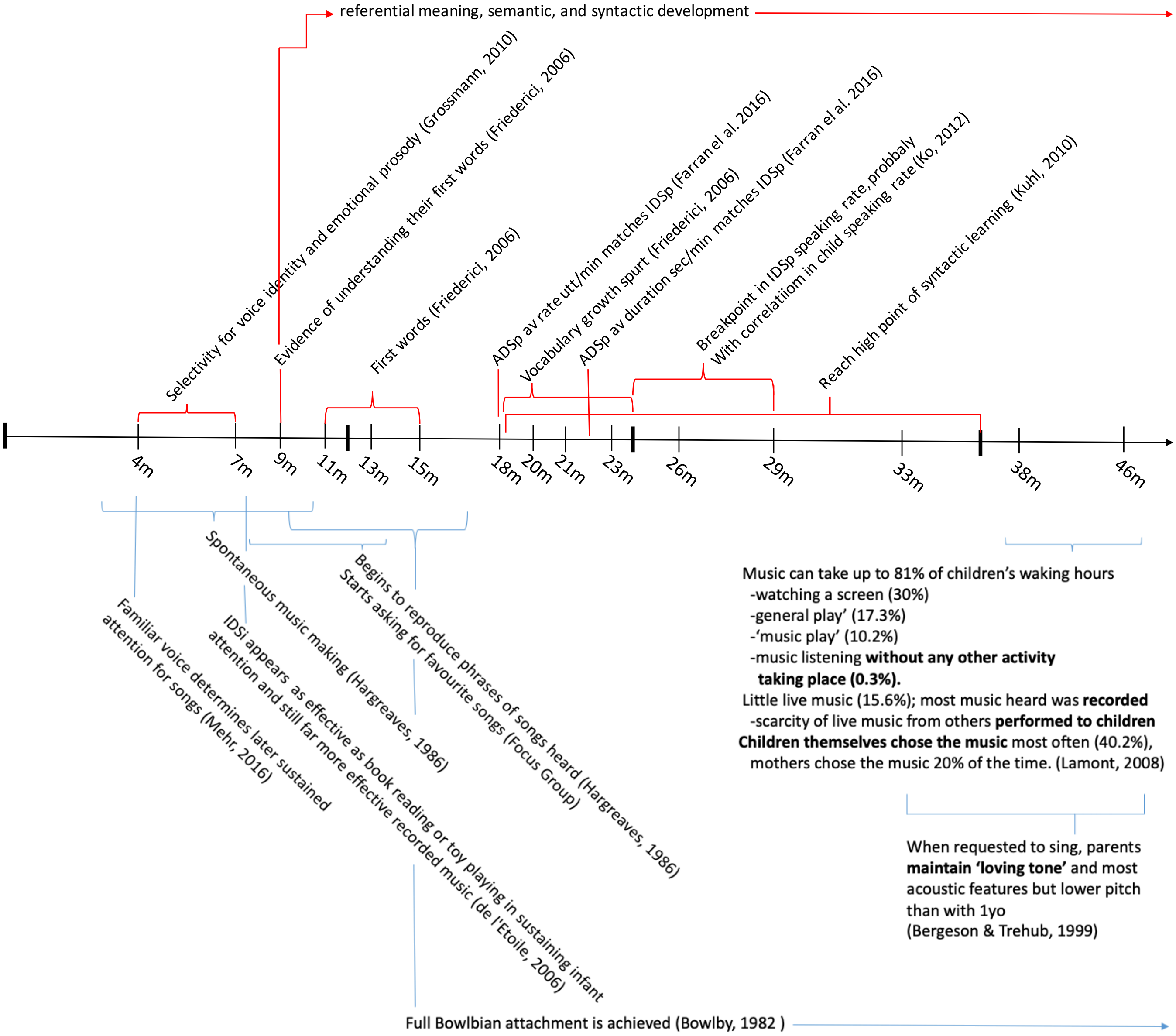


Figure 4. Timeline summarizing the parallel development of infant attention to music (below) and speech (above), the latter in and outside the IDReg.

Before beginning any empirical assessment of the main thesis, one important caveat must be acknowledged. Ko, Vosoughi and Roy, and Farran's results share one important limitation that directly concerns VAW: all of their data remain as partial measurements of IDSp, and do not directly inform us about affective quality of the latter. As a consequence, it remains entirely unclear whether the reported changes in f0 or speech rate are accompanied or not by changes in the affectionate quality that characterizes IDSp. In other words, it is possible that the systematic changes in IDSp reported by Ko, Vosoughi and Roy, and Farran take place without affecting the accompanying 'loving tone', and that the latter remains constant during our chosen developmental window, only to change later on in the child's development. Nevertheless, the theoretical reasons laid out throughout this chapter make it sensible for us to still consider the possibility that the baseline amount of affection in IDSp will change along with other prosodic features. Qualitative assessments would be needed in order to corroborate that changes in raw measures such as f0 mean and speech rate are effectively paralleled by changes in the 'loving tone' that characterizes IDSp.

Equally importantly, the aforementioned studies do not consider IDSi. A remaining question then is what happens with the use of IDSi around the second year of life. Does it also diminish significantly? In this respect, many possible scenarios emerge. First, it could be perfectly possible that the use of IDSi decreases along with IDSp as part of a general tendency, making music from other sources a main socio-affective compensatory measure. Second, the use of IDSi could remain unaltered by the decrease in IDSp, becoming a privileged source of intensified affect. Third, as a consequence of this second alternative, because IDSp decreases, the use of IDSi could be somehow intensified. In order to asses this, it would be necessary to assess the everyday musical environment of children 18-24 months-old, either through constant recording (as Ko (2012) and Vosoughi & Roy (2012) did) or through everyday sampling methodology such as the one used by Lamont (2008). In any case, before that, I will proceed by holding focus groups with parents of children roughly two years-old.

3.5. Musical activities in the developmental window (what to expect)

Having defined a critical window, and before proceeding to lay out the first concrete steps of a research programme that will be constitute the second half of this dissertation, let us provide a brief account of the scope and emphasis that previous literature has given to the relationship that children this age typically have with music.

The three upcoming empirical studies will revolve around children going through the developmental stage of toddlerhood (roughly 18–36 months of age), a period the complexity of which goes well beyond the linguistic sphere. Toddlerhood is a phase of revolution in many senses, as children master walking, come to establish linguistic communication with others, and strive both for making sense of the surrounding environment and gaining independence (Sole, 2016). Indeed, toddlerhood hosts both the prelude and the epicentre of 'the terrible twos', a developmental period widely perceived as a time where control is central at different levels; it is on the one hand disputed between caregiver and toddler (Gallacher, 2005) while the toddler simultaneously struggles to control its own body and behaviour (Green, 2000).

Its focusing on listening to recorded music gives the present dissertation something of a pioneering quality. Already at a general level, a tendency in the literature on early musical behaviour has been reported, to focusing on infancy and the school-aged periods, thus leaving toddlers (and children up to the pre-school years) comparatively-less studied (Barrett, 2010). When it comes to children in our targeted developmental window, rather far from autonomous listening to recorded music, literature seems to evidence a focus on musical interaction and production. The list of parent-child musical activities in toddlerhood goes well beyond listening to music from devices and includes joint and supported singing, creating songs to accompany daily routines or 'making them special' (Custodero, 2006), playing basic instruments, and dancing (Williams et al., 2015). Many of these activities can be thought of as 'musical play'; self-initiated activities that are enjoyable, intrinsically motivated, and led and controlled by children themselves such as vocal play, musical play with instruments and sound objects, and spontaneous movements (Marsh et al., 2016).

Many of these forms of engagement with music, including musical play, take place alongside companions, and therefore con be conceived from an interactional, social perspective in which moving with others is central (Ilari, 2016). As Ilari notices, in this context of cultural learning and social cognition, embodiment is thought to be central, as social cognition is largely based on bodily interaction with the environment (Gabbard, 2014). Accordingly, infants and toddlers are thought to seek out music to a large extent because it engages them not just emotionally and cognitively, but also physically (Custodero, 2002). Such an interactional, embodied perspective fits naturally amid contents already reviewed in chapter 4 (section 1.4) concerning interpersonal synchrony. In other words, moving with others in musical play or other forms of musical engagement relies on the broader human capacity to relate to others in and through synchronised joint action.

The role of interaction— including synchronous interaction —seems to be prevalent and important, which makes sense given its personally-rewarding (affiliative, attachment) quality. Nevertheless, the present dissertation's scope will shift the focus from the relationship between bodily interaction and music, towards solitary listening. Far from denying the prevalence and importance of interpersonal synchrony, this work rather examines how an apparently less social activity such as autonomous music listening takes place— considering that, eventually, this solitary activity will often take over in terms of prevalence. By this token, I will actually focus on children's attention to music beyond interaction and test a possible role of vocal signals in it.

3.6. Methods overview

I have by now finally defined this dissertation's research objects: I will test the influence of the use of the IDReg on 18-24 month-old children's autonomous, sustained attention to recorded music. In this section I will offer a panoramic view of exactly how I intend to test the main thesis, aiming at each step to justify the choices made. Precisely because of it being an overview, in this section I will not go into the level of detail that will be found in the chapters corresponding to the studies themselves. Instead, a panoramic standpoint will be privileged in order for the reader to easily navigate through the three empirical studies and their interrelation.

Throughout the text, an effort has been made to illustrate how and to what extent the research questions that motivate the present work differ from pre-existing ones. While this may provide originality, it also adds difficulty to the task, as fewer empirical works will have directly addressed the questions, and therefore little or no data might be available to rely on. This has already been the case. When wondering about changes in the use of IDSp, we learned that the matter had been addressed rather tangentially. I thus found myself in need of coining the term VAW, and drawing upon leads from different, largely unrelated research programmes.

I thus opted for generating a research programme that would continue to make the most of available clues in the literature. This meant taking some leaps of faith and trusting that what had been previously reported in the literature would continue to be the case in new assessments. Perhaps another way to put things is that the research programme aimed to save time and energy first by not controlling some variables, and to include them in later studies only if previous designs prove not to be informative-enough. I will come back to this idea.

Because this was my first time working with children (all of my previous research experience revolved around adult communicative interaction), my contact with the target population (children 18-24 months-old) remained largely theoretical. Because of this, I—rightly —felt I should proceed carefully when aiming to spend time and money into empirically testing my own ideas. In particular, I felt the need to contrast the impressions that literature reviews and critical thinking had made on me against those of the actual target population. Of course, because children 18-24 months-old cannot be interviewed, their parents were the next desirable choice. For these reasons, I felt compelled to start with a fully qualitative study: a focus group. I thought that parents would show quick, 'gut-feeling' reactions to some of the ideas expressed in this chapter, and that experiencing such reactions would in turn give me 'gut' insight into my own ideas. At the same time, I bore in mind that the subtlety or complexity of some phenomena may well escape parent's capacity for self-assessment. Obtaining parents' impressions concerning IDSp and children's interest in music would, firstly, allow me to have a first sense of the utility of the main hypothesis. Second, testimonies would provide a first impression regarding whether changes in raw measures such as f0 mean and speech rate (as reported in section 3.3.2) are effectively paralleled by changes in the 'loving tone' that characterizes IDSp.

As will be discussed at the end of the next chapter, testimonies from the focus group will turn out not to suggest any decrease in the use of the 'loving tone', nor any particular connection between a decrease in positive affect and children's interest in music. Nevertheless, as mentioned, the possibility that VAW and related phenomena could simply have escaped the parent's self-assessment and recollection could not be excluded. Therefore, a subsequent empirical, quantitative study was conceived in order to see to what extent it would corroborate the focus group's qualitative findings. This second study needed to provide a direct window into 18-24 months-old children's everyday musical environments; the natural milieu where recorded music takes place. To this end, two alternative approaches were contemplated. A first possibility was to 'wire' toddlers, as Vosoughi & Roy (2012) did.

This possibility was very appealing, given the volume of data it returns (several hours per day). Conversations were held between myself, my supervisor and a third professor about the possibility of purchasing a LENA (2020) system, comprising both special microphones and an intelligent software capable of distinguishing speech from noise, making data pre-processing virtually automatic. Although time and effort were directed towards finding funding for such equipment, they were unfortunately not successful. This lack of supplementary funding made alternative non-intelligent recording options impossible due to the enormous data-processing time they entailed. Given the time 'lost' trying to fund the LENA initiative, I felt compelled to plunge immediately into the next alternative: sampling methodology such as the one reported in Lamont's (2008) study.

The aspect that rendered the experience sampling alternative comparatively-less suitable is that it does not contemplate recording the parent's voice— and therefore their potential use of IDSp. I thus decided to take a leap of faith, and trust Farran's (2016) and other's reports, according to which parents would be diminishing their use of IDSp during this particular phase of development of their children. In other words, I decided not to directly assess the parental speech variable. Notwithstanding, the experience sampling methodology envisaged a rich list of variables, particularly those concerning the children's environment (e.g. the time of the day, the current activity, etc). Notably, the methodology could also be used to assess parental use of IDSi. An effort was thus made to integrate as many variables as possible in the analyses. This was not a simple task, not merely because of the number of variables, but because of their often-differing nature. While the target (or dependent) variable was a non-numerical value (a categorical, or qualitative output: whether children seem to be paying sustained attention to music at the moment of the call, or not), other were ordinal (e.g. genre of the music being played) or quantitative (e.g. time of the day). As a result, a non-linear and complex relationship between the features and the output variable could be foreseen, a scenario in which most classical approaches (e.g. linear or logistic regressions) often fail to return informative outputs. In this context, I was advised to make use of decision tree learning, a simple form of machine learning that outperforms classical approaches when it comes to handling many, dissimilar variables. Decision trees are also convenient insofar they are capable of great predictive power. The exact logic and scope of this statistical tool will be detailed in the second study's methods section.

The second study results suggested quite strongly that the main factor behind sustained attention to recorded music was ongoing interaction with a significant other. Much like those of the first study, these results will be shown not to support the idea that the linguistic sphere had any direct impact on musical engagement with recorded music. Beyond a certain disappointment, I drew the following methodological conclusions. The two previous studies had concurred in not returning anything resembling VAW, therefore— and although the idea still made sense to me —suggesting that VAW was at least not taking place in the assessed developmental window. Nevertheless, such a conclusion could not be reached without first assessing directly all factors involved. In other words, I had taken leap a of faith, hoping for linguistic changes to take place and have an impact on parental use of IDSp as suggested by previous literature. Because data did not seem to evidence an impact of linguistic development on music listening, the next step was to obtain direct measures of parental speech, though time. In a similar vein, the second study was less efficient insofar it did not focus enough on the target phenomenon. Indeed, experience sampling offered a window into the children's environment in general, leaving to chance access to instances of listening to recorded music. Therefore, it seemed necessary to devise a next study that would maximize the chances of observing children's reactions to recorded music, *without* interpersonal interaction priming such reactions.

A third study was thus conceived, meant to rely less on existing literature than the previous ones, instead directly assessing the temporal unfolding of all relevant variables. Envisaging both qualitative and quantitative data, the third study constituted a compromise between the first and second studies. Having found it much harder than I thought to recruit participants for the second study without any monetary incentives, the prospect of losing once again substantial amounts of time seemed undesirable. Such a prospect seemed, unfortunately, even more likely than in the case of the second study given that the present project required sustained, months-long commitment. For these reasons, and fearing an even bigger bottleneck effect, this time I decided to offer monetary incentives. Obtaining funding was not easy— more than one body turned my applications down, and only my faculty provided— partial —support. As generous as it was, this funding was not enough to cover the recruitment expenses. I decided to pay for the remaining gap myself, a decision I certainly do not regret.

The rather tight financial component meant that I could only recruit a handful of participants as case studies. The promise of this choice consisted instead in the opportunity to directly witness my research object and analyse it in depth. I appreciated the ecological validity of the experience sampling methodology and wished to assess children in their daily milieu. For this reason, I aimed to design a distal form of assessment: the one-way remote video recording technique (ORViRT). An additional reason for designing a remote assessment protocol involved— once again —recruitment. A further strategy to find participants was to extend the geographical recruitment area and seek for participants beyond the Cambridge area. This proved useful, as most participants turned out not to reside in Cambridgeshire, and their participation in the study was made possible by the ORViRT. Since the number of case studies was already not the study's forte, I decided to split cases into two sub-samples, one starting at 15 months of age, and another at 23. The idea was to cover the lower and upper limits of the selected developmental window, with children both entering and outgrowing the latter during the study. As previously mentioned, this last study's main strength consisted in the opportunity to witness and analyse in depth both children's attention to recorded music as well as their caregivers' prosody. Accordingly, data obtained from each case was analysed both separately and in an aggregated manner, qualitatively and quantitatively, and finally integrated and discussed. The final product of this third study is a thoroughly informative assessment of the research question.

CHAPTER 6. FIRST STUDY: FOCUS GROUP. WHAT DO PARENTS THINK?

1. Introduction

Before designing a study that would directly assess our hypothesis, it was considered important to contrast some of the main theoretical notions laid out in the previous chapter with the direct experience of the characters that actually perform IDSp: parents. The premise was that circulating some of the main ideas with our target population would allow them to share their own appraisal of such ideas, based on their own account of raising children.

At a general level, obtaining parents' impressions of IDSp and children's interest in music would allow us to roughly test the sensibility of our main hypothesis. More specifically, a qualitative assessment was deemed necessary in order to corroborate that changes in raw measures such as f0 mean and speech rate are effectively paralleled by changes in the 'loving tone' that characterizes IDSp. A focus group was thus established in order to obtain this qualitative assessment.

The aim of the focus group was to obtain the perceptions, opinions, beliefs, and attitudes of caregivers towards their use of the IDReg, and to assess their own impressions of a potential progressive discarding of IDReg as children grow older. By this token, this study will serve as a first test to our thesis and guide a first quantitative study.

2. Methods

2.1 Participants

Participants were previously contacted and invited to participate through e-flyers, posters and leaflets distributed around Cambridge. They were offered a £12 reward in exchange for a 2-hour session. Due to a certain scarcity of interested parents and last-minute dropouts, the focus group had to be postponed twice. Finally, eight mothers and two fathers participated in the focus group[16]. The group's demographics are summarized in Table 1. The age of the parents ranged from 30- to 50-years-old (mean = 40.8). Their 'target' children (the child in the targeted developmental window) ranged from 1.75- to 2.5-years-old. The child of a married couple of participants was adopted at two weeks of age.

	Average	SD
Age	40.8	8.1
Latest qualification[17]	6.8	0.8
Occupation hours per week	25	18.3
Average daily hours spent with the target child	16	7.6
Total number of children	1.2	0.4
Target Child's age (months)	22.8	2.2

Table 1. Summary of participants data

2.2 Procedure

Participants were welcomed by the researcher and an assistant upon arrival. The researcher— familiar with the subject, research question and the intended use of the focus group findings —acted as facilitator. The assistant handled the recording of the session, acquired the necessary materials, and entertained the children brought to the session by playing with them in the nearby garden. The facilitator handed the participants an information sheet and an ethical consent form previously approved by the music faculty. After these documents had been signed, the facilitator proceeded to an initial briefing of the session that consisted of welcoming and thanking the participants, detailing timings and

---

[16] Due to last-minute withdrawals, the number of fathers was significantly lower than that of mothers. However, as will be acknowledged in the next two studies presented in this work, such a proportion in the participation of both genders (masculine and feminine) proves entirely representative.

[17] Figures in the Latest Qualification row correspond to those stipulated by the British government (https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels): Level 1 (first certificate; GCSE - grades 3, 2, 1 or grades D, E, F, G; national vocational qualification (NVQ), music grades 1, 2 and 3; or equivalent) Level 2 (CSE - grade 1; GCSE - grades 9, 8, 7, 6, 5, 4 or grades A*, A, B, C; intermediate apprenticeship; music grades 4 and 5 O level - grade A, B or C; or equivalent) Level 3 (A level, access to higher education diploma; advanced apprenticeship; international Baccalaureate diploma; tech level; or equivalent) Level 4 (certificate of higher education (CertHE); higher apprenticeship; higher national certificate (HNC); or equivalent) Level 5 (diploma of higher education (DipHE); foundation degree; higher national diploma (HND); or equivalent) Level 6 (degree apprenticeship; degree with(out) honours - for example bachelor of the arts (BA) hons, bachelor of science (BSc) hons; graduate certificate; or equivalent) Level 7 (integrated master's degree, for example master of engineering (MEng); master's degree, for example master of arts (MA), master of science (MSc); postgraduate certificate; or equivalent) Level 8 (doctorate, for example doctor of philosophy (PhD or DPhil); or equivalent).

structure (see Table 2), reminding them of the recording methods, clarifying the roles[18], expectations and responsibilities of both participants and facilitators. Overall, it was explained that questions would be asked in an interactive group setting where participants were free to talk with other group members.

Section	Time
1. Briefing	15'
2. Ice-breaker	10'
3. Conversation	80’
4. Debriefing	15'

Table 2. Structure and timing of the focus group session sections

After the briefing, recording of the session began. A Canon Legria digital camera was positioned behind the facilitator so as to provide an angle as panoramic and encompassing as possible. Sound was also recorded by means of an iPhone placed at the centre of the table which everybody sat around. A short ice-breaker section then took place, where the facilitator and participants introduced themselves and their target children. The main conversation followed the ice-breaker section. The facilitator had prepared a set of questions in order to assess the participants’ perceptions, opinions, beliefs, and attitudes towards the IDReg and related topics (Appendix A). The main topics included: what the IDReg is (whether the participants had any awareness of using it, what it felt like to use it, etc.), what they thought their use of the IDReg depended on (instances, actors involved, etc.), and potential changes throughout time of their use of it. The questions were not asked in a strict order, but in line with topics aimed to be covered by privileging the participants’ spontaneous strings of thoughts and/or shifts of topics (Porta, 2014). Finally, participants were debriefed and offered the opportunity to ask any final questions as well as being provided with a contact sheet.

[18] Clarifying roles is the same as introducing the ‘rules of the game’: mutual respect and refraining from passing judgement on other participants’ opinions and testimonies, etc.

2.3. Analysis

The footage and audio recording of the focus group were used to transcribe the session (see Appendix G). Once transcribed, a thematic analysis was performed (Braun & Clarke, 2006) in order to search the data set (in this case, the focus group) and find repeated patterns related to pre-defined research questions. Each of the participants' sentences was considered as Data Items (an individual piece of data collected within a set) long enough to contain a unit of meaning and potentially containing one or many data extracts— an individual, coded chunk of data (see example 1). Any seemingly relevant data extract was assigned one or more codes, the latter comprising a potential theme and the number of times such a theme had been mentioned throughout the transcription thus far. Extracts of data were coded in as many different 'themes' as they fitted into. As a result, an extract may be un-coded, coded once, or coded many times, as relevant. Themes therefore capture important aspects of the data in relation to the research questions, as they represent some level of patterned response or meaning. Ideally, there will be a number of instances of a given theme across the data set (prevalence), but more instances do not necessarily mean the theme itself is more crucial. Rather, a theme's importance lies in whether it is able to capture something important in relation to the overall research questions. Some initial themes may prove relevant enough to become Main Themes, whereas several others may collapse into each other.

> DF: "The anger makes them scared, even if it's just a word [F6]. IDSp is part of maintaining that safe environment. Language becomes a way of creating a safety net around you, for both of us actually". [PS10 SS4 STT4]

Example 1. A sentence transcript. The capital letters 'DF' represent the participant's unique identifier. The whole sentence stands as a Data Item, in this case containing two Data Extracts, each of them coded. The first Data Extract was coded 'F' to stand for the theme 'Fear', and '6' because of being the 6th time such a theme had so far appeared in the transcription. The second Data Extract bears several codes: Psychological safety (PS), Self-Soothing (SS) and Safe Transactional Tone (STT).

Once again, given that there were specific goals motivating the focus group, a theoretical or deductive (‘top-down’) thematic analysis approach was chosen in order to define a selection of themes (Braun & Clarke, 2006). This, as opposed to an inductive or ‘bottom up’ approach, where the researcher is comparatively-less literature-oriented in the shaping of data structure. At a general level in this study’s case, engagement with the literature concerned areas covered in the first four chapters, such as IDSp, its relationship to music, and its use in regards of children’s development. A at more particular level, analysis sought to extract any testimonies that would confirm, contradict or provide nuance to this dissertation’s main thesis and any of its aspects as structured in chapter 5. The rationale behind having literature firmly present was to enhance the analysis by being sensitized to subtle features of the data (Tuckett, 2005). As presented in Table 3, a total of 49 different themes were assigned to the different Data Extracts. Some of the themes were used at a semantic (or, explicit) level (Boyatzis, 1998), where the theme is literally mentioned by the participant (e.g. the theme ‘dance’ was used whenever parents would literally talk about dancing). Other themes were used at a latent or interpretative level, where the theme is not literally mentioned by the participant but associated by the researcher with phenomena described in the literature. For instance, the theme ‘entrainment’ was attributed to the following sentence

> DF: “If he’s agitated, slowing down the speech and the rhythm seems to calm him down as well because you can just… get back to a rhythm that seems more soothing...”

where temporal entrainment applied by means of speech is used to generate emotional entrainment.

In the case of Main Themes, analysis also sought to organize data so that it would confirm, contradict or provide nuance to this dissertation's main thesis in a manner as straightforward as possible. To this effect, three Main Themes were chosen by the researcher: Quantal changes in IDSp, Vocal-Affective weaning, and Music as a compensatory affective resource against VAW. Following the examiner's advice, a fourth Main Theme, Mentalisation and the Theory of natural pedagogy was added. The next section presents the most important perceptions, opinions, beliefs, and attitudes concerning such Main themes, and illustrates them through relevant excerpts containing one or more Data Items from the session's transcription.

Theme	Abbreviation	Illustrating quote
Attachment	A	*"I think for us, since we adopted Richie, we didn't have that same biological **attachment**"*
Adult inhibition	AI	*"I mean, obviously you're **not gonna show your excitement** like you're a kid"*
Attitudinal prosody	AP	*"sometimes the **IDSp is just to convey** a safe space"*
Authority	Au	*"you can change your tone to be slightly more **authoritative**"*
Audience Effect	AE	*"And occasionally it becomes— embarrassingly enough —the voice you use **in public** then you really want him to do something and **you don't wanna be embarrassed**"*
Communicative compensation	CC	*"I'm probably more relaxed in the sense of not thinking **to switch the tone** and using my normal voice because de understands what I'm saying in terms of the words"*
Code-Switching	CS	*"I don't know why one **does a change**. Maybe like when you **talk to an elderly person** and you try to be a bit more formal"*
Cultural Learning	CL	*"you use also a special kind of tone like when **trying to explain** like something works"*
Danger	D	*"you need to distinguish from when you have to say something like '**don't play in traffic**"*
Dance	Da	*"And then we had a piece that I knew I played it when I was pregnant, and I knew that my children recognized it every time… but **now it's dancing."***
Disinhibition	Des	*"I think there is something **not as inhibiting with your own child**, as there is perhaps with others"*
Dyadic dynamic	DD	*"So, there might be some news that I share with other adults that I'm excited about, but that **level of interaction with your own child** is something that you can sort of just be yourself"*
Entrainment	E	*"…**slowing down the speech and the rhythm seems to calm him down** as well because you can just… get back to a rhythm that seems more soothing."*
Emotional Contagion	EC	*"he won't be calm unless you are calm"*
Emotional Complexity	ECx	*"I think **very rarely an emotion in the outside world is that simple**. You know, you might laugh at a joke but at the same time feel guilty that you laughed at a co-worker and you're being silly"*
Emotional Modulation	EM	*"…**slowing down the speech and the rhythm seems to calm him down** as well because you can just… get back to a rhythm that seems more soothing."*

Empathy	Emp	*"There are these moments like 'he's learning something big right now'* ***and you can feel it"***
Emotional processing	EP	*"I think something similar happened to me… whenever I would try to sooth her* ***or to help processing something"***
Engagement	En	*"So, I think that was definitively powerful for him. Talking in that way to really sort of* ***engage them****, whether I was trying to sooth them or not"*
Eye Contact	EyC	*"it strikes me that I spoke in that way when I was trying to* ***engage in eye contact"***
Fear	F	*"I mean, there are many situations when things have escalated to a very authoritative tone, but it* ***terrifies him****."*
Horizontality	H	*"so, if I'm addressing him, I'm teaching him how to speak, but I think I'm* ***addressing him as a normal person****."*
Intimacy	In	*"It was very good at turning into that particular* ***closeness****, whether I was looking for their attention, or trying to distract them, it was all very sort of a* ***close moment****."*
Irreflexibity	Ir	*"And as such, it* ***doesn't have the same inhibition****. There's* ***no thought for it****, you know."*
Infant-carer conflict	ICC	*"but also when you're trying to get them to behave or not-do something they're trying to do, as in 'hold my hand' or 'don't run into the road' (…) you change your tone when you're trying to* ***get some control"***
I-D disinhibitions	IDD	*"you can equally be silly, or enthusiastic, or whatever it is because they love you whatever you're like so, you* ***get that chance to be no inhibitions****."*
IDReg Demand	IRD	*"This morning he came to me and gave me a huge hug and said 'aw, my hero', and he said it in a babytalk way,* ***so maybe he was asking for it****, and I did reply in that sort of tone"*
IDReg Persistence	IRP	*"whenever I would try to sooth her or to help processing something, that is when it* ***came back****"*
Language acquisition	LA	*"Perhaps I realized over a month… but it's like a quick switch, it's like they suddenly start making sentences"*
Language Barrier	LB	*"So, if I ask 'do you want a yogurt?' he's really frustrated because he's trying to tell me what he wants* ***but he can't"***
Linguistic Power	LP	*"And* ***once the language comes****,* ***he's much more confident in asserting himself*** *(laughs)"*
Mentalisation	Mnt	*"I think something similar happened to me… whenever I would* ***try to sooth her [S5] or to help processing something…"***
Music	M	*"It was one of Haydn's symphonies. But only that one. He knew it straight away and [snores] it was like an off-switch"*
Multimodality	MM	*"It was some Disney animated movies, because there was* ***video as well as sound"***
Music preference	MP	*"Right now, it's the song from Trolls. He wasn't that interested in the movie; all he wants is the song. But when he was smaller, it was the Muppets, then it was Frozen…"*
Music Preference Phase	MPP	*"we've had different songs* ***at different phases****. Right now, it's the song from Trolls."*
Theory of natural Pedagogy [ToNp]	NP	*"[…] you use also a special kind of tone like when trying to explain the way something works."*
Positive Emotion Preponderance	PEP	*"And I think that's really nice, so I think there's a degree of expressiveness that is* ***much broader on the silly, fun side*** *with him, than on the serious, frustration and all of that other stuff…"*
Prenatal Music	PM	*"And then we had a piece that I knew I* ***played it when I was pregnant,*** *and* ***I knew that my children recognized it every time…"***
Personality [identity] Plenitude	PP	*"So, there might be some news that I share with other adults that I'm excited about, but that level of interaction with your own child is something that* ***you can sort of just be yourself"***
Perceived safety	PS	*"I think the speech then became very much for us about providing that* ***sense of safety*** *for him"*
Pedomorphic traits	PT	*"the infant-directed, the soothing, the bringing things down a level in a way, and the authoritative edge, like 'I'm serious now' and it's hard*

		because he's cute"
Social Rules	SR	*"That change in the voice from sweeter and more child-like, to one that conveys that there is something that **shouldn't be done"***
Self-soothing	SS	*"Because when he's getting agitated you don't wanna get agitated too, so if you force yourself into that frame of mind, that slowness, **it actually calms both of us down**, in a weird way."*
Social Status	SoS	*"Maybe like when you **talk to an elderly person** and you try to be a bit more formal…"*
Soothing	S	*"So that's a third one, but I think it comes down more to **skill acquisition and soothing** more than anything else now."*
Safe Transactional Tone	STT	*"there's this fear inside them and the **IDSp with the authoritative tone embedded in it**, it's meant to take out the fear"*
Tantrum	T	*"the worst thing if he's upset is for you to be upset and stressed, which is hard if you're having a tantrum"*
Tone Specificity	TS	*"So, when the will comes out, then you have to have more than just two tones, you have to have a **battery of tones**, so that they can understand."*

Table 3. Complete list of themes. Abbreviations are presented as used both in the coding of the original transcription (see Appendix G) and in the excerpts presented in this chapter. Bold font is used to enhance a portion of a quote that was particularly crucial in the attribution of its associated theme.

3. Main Themes

3.1. Quantal changes in IDSp

There was a general awareness of changes in IDSp surrounding language acquisition among participants. Most participants could not ratify a quantal shift but could still give examples for a change in their use of IDSp.

Excerpt 1

Researcher: what I meant to say is that you mentioned that the register would sometimes compensate for the lack of understanding of words. Has that changed now that Richie actually understands words?

DF: yeah, I guess it has. I'm probably more relaxed in the sense of not thinking to switch [CS3 CC2 Ir1] the tone and using my normal voice because he [her son] understands what I'm saying in terms of the words [LB2]. So, I think that makes it a bit easier to just communicate on a day to day basis. I'm from NY, I talk fast, I talk loud (laughs), so slowing myself down is sometimes hard, but, yeah, as he's gained more language, I guess it's less infant-directed and we're starting to have more conversations, so when there's a conversational element, then it becomes much more of a reflexive tone in terms of what's the tone he's using and how is that going, rather than something in which I'm trying to define the tone of the conversation [E2].

Alternatively, some participants seemed to clearly detect a quantal change in their use of IDSp as a result of acknowledging their children acquiring language:

Ex.2

R: and the fact that your daughter talks and you have at some point acknowledged it, has had any impact in the way you address her?

C: maybe yes. I think the fact that she understands makes you shift from child-like to more normal [LB4 SoS3].

J: Still, one keeps using childish expressions like 'oh look', but I do feel like the fact that they understand you means that you can address them more normally [NP7], with the tone you would use when addressing another person.

C: Before, you didn't know whether she was getting was you were transmitting or not [LB5], hence the slower, affectionate, and fragile or delicate tone[NP8]. I can now say "Alondra, put on your jumper because you're going to play outside".

Interestingly, some of them explained it in terms of a quantal change in the 'social status' of their children as a consequence of talking, and a consequent change in the way they address them.

Ex.3

R: I see. You mentioned something more like 'fragile' in her that made you feel like talking to her like that. What do you think would happen if you would address her like any other adult, like I'm talking to you now? What would be the difference?

C: maybe none… (…) but I think that… I don't know why one does a change. Maybe like when you talk to an elderly person and you try to be a bit more formal… [SoS1 CS2]

R: like someone who has a different status?

A: yes, a different status, more serious, who you can't talk to… I think it's not something you think, because they're smaller [SoS2 CS3], and not just with my boy, but with other children, you make that shift in the voice…

3.2 Vocal-Affective weaning

Most participants claimed maintaining a rather steady amount of affection in their vocalization to their children even when not using IDSp.

Ex. 4

> C: it's true that you demand more from them, but at the same time, they still need much attention from you. But at the same time, they would like to do things on their own. Alondra at a stage in which she likes doing things on her own, but still needs much affection.

The main change in the affective quality of the participant's speech after acknowledging their target children acquiring language seems to be an increasing amount of anger, a consequence of the increased Infant-Caregiver Conflict that language acquisition implies. Participants explained that a certain degree of 'Linguistic Power' is attached to language: further resources for the child in terms of fulfilling its will.

Ex.5

> R: Is this related to, for instance, to the 'qualitative status shift' Coline previously referred to?
>
> DF: no, I think it's true, I think the will comes up before the language, as Grace was saying, I think the will you can start to see before, but I thought the will before the language was like almost tentative, in the sense that he was trying to see 'could I assert myself?' And once the language comes, he's much more confident in asserting himself [LAX LP2] (laughs), so you can see, in a way, it's not just the 'adult emotions' I think it's the growing confidence of his control over his own environment [LP3]. (…) you know, it's that sense of increased confidence in his control of the environment [LP5], that he can do things to get the reaction that he wants. And he'll assess to what degree he can manipulate you…

Thus, a talking child makes uncooperativeness and conflict evident:

Ex.6

C: it's true that once they've learned to talk you demand more from them, but at the same time, they still need much attention from one. But at the same time, they would like to do things on their own. Alondra is at a stage in which she likes doing things on her own, but still needs much affection. For instance, I ask her to tidy up. One expects them to cooperate, but at the same time when asking for something, there is a certain rejection, and a clash [ICC4]. You expect to be understood and helped, but at the same time there is a quarrel in which she is not responding. It's not because she doesn't understand, but because she doesn't want to. I think it's a natural phase of rebelliousness. It's tricky. Sometimes you have to be firmer [Au4].

R: so, since she talks, is there more conflict?

C: yes, absolutely. She understands and understands that I understand. We both know we understand each other. But she might simply not want to cooperate.

In this context, participants' avoidance of negative emotions was made clear. IDSp was presented as a means for minimising negative emotional prosody— anger and panic in particular —thus avoiding scaring the child and maintaining a 'safe Psychological Environment' where overt anger is— ideally —only reserved for situations of imminent danger.

Ex.7

DF: and safely, because… I mean, there are many situations when things have escalated to a very authoritative tone, but it terrifies him. I mean, I terrifies him! [AP5 PS6 F2]

G: yeah, it's about not getting him scared [PS7 F3]

DS: that's actually about trying to… because you have a range [TS4] from 'I'm asking nicely', to 'I'm asking a bit more urgently', to 'we really need to do it now', and the top one is… once he nearly run down into the road and I screamed 'no!' at the top of my lungs and it scared the crap out of him… but he stopped.

(…)

DF: "The anger makes them scared, even if it's just a word [F6]. IDSp is part of maintaining that safe environment. Language becomes a way of creating a safety net around you, for both of us actually". [PS10 SS4 STT4]

IDSp was also reported to help maintaining such a 'safe Psychological Environment' not simply by influencing the addressee, but as part of a self-soothing effort of the parents.

Ex.8

P: it so strikes me that you're trying to share them being in the same wave-length as you [E4 I?], kind of bearing in mind what would be beneficial if they were… But you know that if they're becoming very aroused, and you want them to get back to the baseline [EM5], so it's like you take a deep breath, and you feel calm [SS2] and then by speaking to them… I think it's more than speech, it's a lot to do with your household-self [MM1], but speech is part of it, and that helps to regulate them back to that level that you hope it's the same wave-length that you're on… [E5 EC4 I?] hopefully (laughs)"

Even further, sometimes IDSp would have a psychological effect in the parent itself as he or she would use it:

Ex.9

R: Why do you use it [IDSp] when you feel like using it?

DF: usually to keep him [Richie] calm. If he's agitated, slowing down the speech and the rhythm seems to calm him down as well because you can just… get back to a rhythm that seems more soothing [I1 S1 E1 EM1]. But I think it also sooths me, selfishly enough (chuckles). Because when he's getting agitated you don't wanna' get agitated too, so if you force yourself into that frame of mind, that slowness, it actually calms both of us down, in a weird way. [I2 S2 E2 SS1 EM1]

IDSp thus seems to allow the expression of a degree of anger that does not threat interpersonal bonds:

Ex.10

DS: that's actually the about trying to… because you have a range [TS4] from 'I'm asking nicely', to 'I'm asking a bit more urgently', to 'we really need to do it now', and the top one is… once he nearly run down into the road and I screamed 'no!' at the top of my lungs and it scared the crap out of him… but he stopped. [D3 A2 F4 LP7 PS7]

DF: you have to use that one so sparingly for it to mean anything [overt anger] [TS5]. Cause they're still young enough that although they're attached, there's still this fear… [A4 F5] there's this fear inside them and the IDSp with the authoritative tone embedded in it it's meant to take out the fear [AP6 PS8 PEP4 STT2].

This "IDSp with the authoritative tone embedded" could be provisionally referred to as 'safe transactional tone'. Conversely, typical IDSp interactions— both before and after language acquisition —seem to concentrate around the 'positive' side of the emotional spectrum in an uninhibited manner that characterizes them.

Ex.11

R: do you feel as excited elsewhere as you would with your children? [explains disinhibited quality of IDSp]

G: I try to. I mean, obviously [SR] you're not gonna show your excitement like you're a kid [AI1], but when something makes me passionate, I try to show excitement.

P: I think there is something not as inhibiting with your own child [IDD1], as there is perhaps with others [AI2]. So, there might be some news that I share with other adults that I'm excited about, but that level of interaction with your own child is something that you can sort of just be yourself… [PP1 IDD2 PS5]

G: exactly [AI3 IDD3]

P: because, they love you however you are [UL1 A2] and you can equally be silly, or enthusiastic, or whatever it is because they love you whatever you're like so, you get that chance to be no inhibitions. [IDD4]

DF: definitively there is a joy that you can express [PEP1] that is just pure [PP2 ECx1]. It's just purer when you're with your child than when you're laughing at a joke at work or with your friends, it doesn't have the same innocence, in a weird way. And as such, it doesn't have the same inhibition [IDD 6]. There's no thought for it [Ir2], you know. And I think that's really nice, so I think there's a degree of expressiveness that is much broader on the silly, fun side [PEP2] with him, than on the serious, frustration and all of that other stuff…

R: [to G, DS, A and M]: anything to add guys?

G: but, for example, now that I get more clearly the example in my head… I'm a very open person, I'm a very sharable person, so I try to express myself, actually, but my husband is exactly the opposite. So, that's why I see the difference between the way he talks to him and the way I talk to him, which is exactly the same outside actually. When he talks to people, he tries to be really serious [AE3 SR2?], really like 'I'm an important person' and I'm more like 'yeah….'

A: I agree… [nodding]

All participants acknowledged their children seeking for IDSp after language acquisition in a way that seemed unambiguous to them.

Ex.12

R: do the introduction of language and the waning of IDSp imply that children sometimes ask for it?

DS: I think he does, in a way… This morning he came to me and gave me a huge hug and said 'aw, my hero', and he said it in a babytalk way, so maybe he was asking for it, and I did reply in that sort of tone [IRD1]

G: my son is quite independent, but when he's tired or bored, he will call me by his side 'come, mom, come', and take my hand and ask me to sit, so that you know that he really wants you to stay with him, sit (…) when he really needs it, he knows how to ask for that kind of speech [IRD2]

R: but isn't it simply about getting attention? Or do you feel he also requires de IDSp?

G: yeah, yeah, because he tells you 'come, mummy' and he starts telling you something that will lead to that kind of talking

DF: I think it's… it's exactly that, it's more than just attention, there are times when he wants that soothing-ness [IRD3], and you can tell when it's not just 'I don't want you to pay attention to something else' but it's that soothing-ness that he needs. And to Dean's point, he'll bring out a bit of a baby voice, or bring out more of a [pout] [IRD4]

R: and in which kind of instances would they need you like that?

G: when I realise it the more is when he's tired, when he wants to go to sleep. So, he's like 'come one, mummy', and he knows that when we're at the sofa it's like 'relaxing time', and he's knows it [IRD5]

DF: I get it, directed at me, at least, when there's a change in the routine [IRD6 A3]. For instance, I usually wake him up and then come back home at 6, but last week I had to stay in London Thursday night, and when I got back, he was clingy for the whole weekend. The power of routine at this age is so huge, and the moment I lose any of my 'assigned roles' in that routine [laughs] he becomes much more focused on the soothing than anything else [IRD7].

3.3. Music as a compensatory affective resource against VAW

No direct connection between music and VAW was reported by participants. The only explicit change concerning their children's musical life related to language acquisition was their gaining the ability for distinguishing and recognizing musical instruments. There were some examples of preference for instrumental music before language acquisition.

Ex.13

R: when did you realize she had favourite songs, or that there where songs she remembers?

C: Twinkle-twinkle was some five months ago [1.5 yo], more or less when she could talk. [MP2 LA3]

R: did she la la la the songs before that?

C: could be… can't remember. She would ask me for songs [MP3], whose lyrics she couldn't really remember. I do remember more lately, now that she sings.

R: and before any singing and la la la, did you noticed any particular preference for a given piece of music?

C: yes, around the first year [MP4]. It was some Disney animated movies, because there was video as well as sound [MM1].

However, the preference stated in Excerpt 11 seems probably more related to the multimodality of the stimulus than to music itself. Different is the following case:

Ex.14

> DS*: when he was really little, like 3 or 4 months old, there was one piece I could put on and he'd be asleep like in 3 mins. It was one of Haydn's symphonies. But only that one. He knew it straight away and [snores] it was like an off-switch* [M2 A5 S6 MP2]

This second example suggests that a non-vocal, instrumental piece of music had a stable effect on the infant. Rather than the infant having an affective motivation for listening to the Haydn piece, it can be said that the piece had a stable effect on it (e.g. possible 'familiarity effect') which doesn't explain how it starts being effective, but only how it comes to be effective.

There also were mentions of preference for music heard during gestation:

Ex.15

> P: (…) there were many pieces of music that we enjoyed over the years [MPP2]. And then we had a piece that I knew I played it when I was pregnant, and I knew that my children recognized it every time…
>
> R: well, we talked about this sort of 'safe space'…
>
> G: as I said, he prefers French music over the lullabies at the nursery because that's the music he heard while I was pregnant [PM2] and during the first months of being born in France. And that's the music he'd prefer to sing. [MP3]

Finally, there were mentions of dancing as a means for dyadic interaction:

Ex.16

DS: when he was really little, like 3 or 4 months old, there was one piece I could put on and he'd be asleep like in 3 mins. It was one of Haydn's symphonies. But only that one. He knew it straight away and [snores] it was like an off-switch [M2 A5 S6 MP2]

R: so, he had like a 'favourite song' from very early on…

DF: so, there were a number of songs in the sense of… music that we could put on that would change the mood… communicate something [M3 EM7]. Right now Trolls is just a dancing thing… [Da1]

DS: yeah, but it's something between you and him… [DD4]

P: as you were saying this, I was thinking that we sort of moved from a stage where we had singing classes and we sung to the children and there were many pieces of music that we enjoyed over the years [MPP2]. And then we had a piece that I knew I played it when I was pregnant, and I knew that my children recognized it every time… [PM1] but now it's dancing [Da2]. The music is still involved very much, but it's just that way of sort unwinding down at the end of the day… you know, if you're a bit tired or it's been a hectic day or it's been a bit stressful, we dance in the kitchen. I don't even care if people could see me… [Des1]

3.4. Mentalisation and natural pedagogy

Participating parents narrated episodes where they made efforts to interpret the children's behaviour as meaningful clues to their inner experience— in other words, mentalisation. For example, more than one narration depicted the role of mentalisation in assisting children whose linguistic abilities would not allow them to fully express their needs or desires:

Ex. 17

M: in my case, although she cannot yet put a sentence together, she would go [bumbles babbling in a speech-like manner] and at the and say 'foot', so that you know her foot is involved in something else [Mnt7]

Ex.18

G: exactly [the latter]. For instance, the 'yes' and 'no' he understands perfectly. So, if I ask 'do you want a yogurt?' he's really frustrated because he's trying to tell me what he wants but he can't [LB3]. So, I ask him ‘is it a banana you want?’ and he would shake his head, until I ask him what he finally wants, like the milk, and he finally nods in agreement and you give him the milk and he’s happy [Mnt5]. But before that he was so frustrated because he couldn’t express himself, but I know he’s actually understanding what I’m saying [Mnt6].

These examples thus depict how attentive caregivers perceive and infer their children’s mental processes, which in turn enables them to fill in the content gaps of half-babbled sentences (excerpt 17) or to assess the child’s degree of understanding of their own adult behaviour (excerpt 18). Sometimes, parental mentalisation was used in moments when the child was not able to regulate their own emotions. In excerpt number 8 Paula describes how, when facing her overly-excited child, first regulates her own affective state taking a deep breath. The effort of acknowledging her own anxiety constitutes a first form of mentalisation, as she is looking at herself ‘from the outside’ (see chapter 3 section 1.3). Her self-regulation allows in turn acknowledging her son’s internal state ‘from the inside’, which corresponds to a second aspect of mentalisation. Such a second acknowledgement allows her to, through her use of IDSp, help him regulate his own emotional state— she actually uses the word regulation—, to bring him ‘back to baseline’. Something similar is depicted by DorothyF, who uses IDSp to calm her son and change his ‘rhythm’ (excerpt 9).

More often than not, parental mentalisation was reported to have been deployed in instances of learning or understanding:

Ex. 19

R: in which other stances you still use IDSp with Richie?

DF: there are these moments like ‘he’s learning something big right now’ [CL3] and you can feel it [Emp1], and you can kind of come to it [Mnt2].

Furthermore, parents reported relying on their tone of voice as a resource for explaining or teaching:

Ex. 20

> G: what I was thinking was that when you want them to understand something [En3 CC1 NP3], you use [IDSp] also a special kind of tone like when trying to explain the way something works.
>
> A: I was just realizing that although he's tiny sometimes he just doesn't do something just because he decided not to. I think it's a mixture of things, not just IDSp. Sometimes you may try a simple sentence, and if he doesn't understand, you may change the tone [CCX Mnt4 NP9]

In excerpt number 20, AVA illustrated an example of how mentalisation can modulate the caregiver's use of IDSp. In the testimony, her assessment of her son's degree of understanding of a given sentence may (or may not) prompt the use of IDSp an ostensive cue that supports the transmission of knowledge, as described in the theory of natural pedagogy (ToNP). In the same excerpt, Grace quite literally conveys the idea that an ostensive cue such as IDSp compensates for a novel object's teleological or casual opacity (see chapter 3, section 2.2.2.1). Excerpt number 21 also evidences the close relationship between IDSp and eye contact often needed in situations involving some kind of knowledge transference (Senju & Csibra, 2008).

Ex. 21

> P: it strikes me that I spoke in that way when I was trying to **engage in eye contact** [EyC1] with him [her son, Newton]

4. Discussion

The aim of this focus group was to assess qualitatively the presence of vocal-affective weaning, defined as the process of gradually introducing a human infant to what will be its adult culturally defined standard range of vocal-affective intake. Additionally, it intended to explore any possible connection between VAW and music outside the ID register while taking into account the possibility that music— a culturally-sanctioned source of affect — should constitute a means for filling the gap that this ‘affective weaning’ generates.

Results are straightforwardly informative insofar as they suggest that the hypothesis must be rendered more specific. VAW has so far been defined as

> the process of gradually introducing a human child to what will be its adult, culturally defined baseline intake of vocally-expressed affection.

Such a phrasing implies that what abounds in the IDReg and decreases in VAW is just any kind of emotional vocalization. This claim will not be fully discarded, but will require more accurate description. The following three paragraphs will aim to do so.

Participants' testimonies did ratify to some degree the notion that typical IDSp interactions— both before and after language acquisition —often coincide with vocal expression of emotions on the positive end of the valence spectrum. For instance, joy was reported as occupying an important place. Such joy would often be intensified through disinhibition for two main reasons. First, as a response to the infant's own expression of joy, harnessed into its pedomorphic traits and pedoacoustic vocalization. Second, by the dyadic dynamic of the interaction, through which further possible implications of the encounter are absent (e.g. hurting someone else's feelings, moral apprehensions, etc.). Similarly, IDSp was declared to be used as a soothing resource and for conveying safety, both pleasant feelings.

Nevertheless, participants asserted that IDSp would also be used when trying to avoid vocal expressions of anger. In accordance with these arguments, embedding angry vocalization in the ID register would counteract the threatening effect of the former, in the form of a ‘safe transactional tone’. Indeed, IDSp and IDSi were reported to be important resources for avoiding conflict, fostering cooperation, and maintaining healthy attachment.

This finding confirms the use— or discarding —of ID register as involved in parent-offspring conflict. As previously explained, in parent-offspring conflict, caregivers and infants are expected to disagree on how long the period of parental investment should last, as well as on the amount of parental investment that should be given. The altruistic and egotistic tendencies of the offspring are also disputed, as these tendencies affect other relatives (most often, siblings) (Trivers, 1974). Accordingly, the egotistic impulse would be for a caregiver to express frustration and anger, while investment would take the form of emotional self-regulation, which implies psychological effort. Thus, in the present results, parents' reported avoidance of overt expressions of anger can be considered a form of investment. From this point of view, it is quite evident that IDSp stands as an interactive device that simultaneously allows parents to minimize their investment (IDSp self-soothing capacity lowers the necessary emotional effort) while at the same time influencing the infant to become less demanding (by soothing them). In communicative terms, this is the equivalent of saying that IDSp constitutes in this context a form of management in the sense introduced by Owings and Morton (see Chapter 2, part I, section 1.1.2), involving self-interested efforts to change the balance in current conflict by regulating the behaviour of their children by means of a signal— the IDReg.

In the same direction, participant's testimonies did not support the idea that language acquisition palpably triggers a decrease in the 'vocal-affective load' of their vocalizations. Instead, testimonies consistently reported an increase in the expression of negative emotions such as anger as a direct consequence of acknowledging language acquisition. Such expression of anger would itself be a result of the increased infant-caregiver conflict that language acquisition implies (e.g. explicit uncooperativeness, rebelliousness. disobedience, etc.). In sum, it seems more accurate to argue that IDSp reflects free vocal expression of emotions that do not directly threaten infants, as opposed to the more inhibited expression of positive affect and the more overt expression of anger that seem to accompany instances of switching to ADSp.

It also seems more accurate to say that IDSp is fundamentally compatible with emotions on the positive side of the valence spectrum, while being comparatively-less so with emotions on the opposite end such as panic, but especially anger. This notion allows us to provide nuance into insights contributed by previous research. For instance, it was suggested that an important function of IDSp concerns the communication of affect (Fernald, 1989, 1991). This kind of speech would render an exaggerated indication of the speaker' affective state, consequently allowing the child to identify emotions more easily. Similarly, it has been argued that IDSp reflects free vocal expression of emotion to infants, as opposed to the more inhibited expression of affect that characterizes most stances of adult-directed speech (Trainor et al., 2000). These two, closely related ideas now appear as rather only partially correct. Trainor focused on the four emotions love, comfort, surprise, and fear, as these are emotions that caregivers are likely to express to infants. In the case of fear, her initial results seem to be at odds with the present ones. One of the participants narrated how he had 'scared the crap' out of his son by screaming 'no!' at the top of his lungs when the child nearly ran down into a road. It is, indeed, not hard to imagine how the tone and loudness of panic would be incompatible with the IDReg. It can thus be argued that the intensity in the expressions of fear of Trainor's subjects was rather subdued, and that differences can be found between expressions of fear and panic (as extreme fear) embedded— or not —in the IDReg. What can be safely speculated is that differences would be found in the acoustic parameters of the vocal expression of anger depending on its degree of embedding in the IDReg.

Another possible dialogue with literature concerns the relationship between emotion and register, as problematized by Singh et al. (2002). As discussed in chapter 4 (section 3.2), the authors experimentally tested infants' listening preference by independently and systematically manipulating affect (happy, neutral, or sad) and speech register (ID v/s AD). As a result, situations were created where 'neutral IDSp' and 'sad IDSp' were directed at participating children, something— as stressed by the authors —not often done in this field of research. While such devised, novel experimental combinations of affect and register indeed illustrated the researcher's point— a confound between these two factors —, they also may have entailed a degree of artificiality. According to the testimonies hereby reported, instances of 'IDSp fear' or 'IDSp anger' are likely to be rather rare when compared to IDSp renderings of positive-valence emotions. Furthermore, more affectively-extreme possibilities such as 'IDSp panic' seem improbable in ecological contexts. Something similar can be said about 'neutral IDSp', a notion that seems to challenge some of the quintessential features of the IDReg— notably, prosodic exaggeration. In sum, although the number of logically-possible combinations between register and emotion is larger than often considered, the probability of their manifestation in everyday life varies importantly. While 'IDSp happy' utterances might have been privileged in the literature, such a choice might not be over-representative, after all.

At this point, the above arguments allow for a reconsideration of the original formulations. As previously mentioned, the original formulation of Vocal-Affective Weaning, initially defined as the process of gradually introducing a human infant to what will be its culturally defined standard range of vocal-affective intake in adulthood, turns out to inaccurate. If the term 'weaning' is to be kept, VAW ought then to be defined as

> the process of gradually introducing a human child to what will be its adult, culturally defined baseline intake of vocally-expressed affection, as well as negative vocal-affective intake— particularly anger.

This redefinition does not rule out, but rather enriches the original formulation. An increase in the expression of negative emotions would equally act as a motivation for children seeking positive emotions, such as those which proved to be compatible with the IDReg. However, it does have a provisional impact on the second goal of this study, which is to explore any possible connection between VAW and music beyond the IDReg. It must also be taken into account that music— a culturally-sanctioned source of affect —may constitute a means for filling the gap that WAV would generate.

As discussed, the present results strongly suggest that the most evident consequence of language acquisition is an increase in the expression of negative emotions such as anger, rather than a decrease in positive emotions. This confirms the idea, presented in the last chapter, that VAW is part of a larger process of enculturation, in which language plays a central role in the progressive introduction of norms and the potentially resulting conflict. If the vocal expression of positive emotions remains rather constant following language acquisition, then there would be no case for music outside the IDReg constituting a means for filling the gap that WAV would generate. This may also explain why participants did not report any particular demand for music in children following language acquisition that was not already present. Many different reasons should be kept in mind.

First, the current sample and method must be scrutinized. It can be hypothesised that the motivation for engaging in such musical dynamics augments with the mentioned increase in negative emotions, but such an augmentation cannot be assessed through a focus group. In other words, in the same manner that some participants were more aware than others of the 'quantal change' in IDSp that Ko (2012) and Farran (2016) report, the presented sample might simply not be aware of an increment in musical interest. Similarly, it might be uncomfortable for parents to acknowledge a decrease in the level of affect in their vocalizations and might thus avoid accepting and/or reporting it.

A second, possible reason to be kept in mind is VAW. Results strongly suggesting that the most evident consequence of language acquisition is an increase in the expression of anger by no means deny the fact that the vocal expression of positive emotions may later wane in a quantal manner. Thus, it might be that it is not around the second year, but later, that music constitutes a means for filling the gap that WAV implies. The challenge then would be finding another moment in development where the waning of positive emotional vocalization occurs in a circumscribed, 'quantal' manner. The possibility that the waning of positive emotional vocalization never occurs in a temporally-circumscribed, 'quantal' manner, cannot be ruled out. In the same line, it must be born in mind that an increment in musical interest might stretch over many years, in such a way that parents cannot perceive it.

Third, the development of music cognition also must be considered. Children's ability to discriminate and label happy versus sad musical excerpts from real music contrasting in tempo and mode is a skill that only improves with time. 3- to 4-year-olds are unable to distinguish them on any basis, 4-5-year-olds can do but only based on tempo and with the aid of visual cues, while 6-8 -year-olds reach adult-like proficiency (Dalla bella et al., 2001; Mote 2011). At the same time, children between the second and third year of life can easily sense the difference between positive and negative emotions. There is a mismatch between the learning curve of emotion recognition in speech and music. It could then be that it is only between ages 3 and 4 that infants start decoding positive emotion in music and are thus able to fill the gap that WAV implies. In any case, arguments in this paragraph only apply when considering that it is the affect expressed through music that mainly draws children's attention. The possibility that other social and/or cognitive factors (linguistic learning, enculturation, etc.) might explain the appeal of music must not be discarded yet. Accordingly, this subject will be discussed in the remaining studies.

Escaping the scope of the main thesis, mentalisation and the theory of natural pedagogy (ToNP) applied to an important number of testimonies. The use of IDSp as an ostensive cue was particularly explicit, and seemed to be more or less consciously acknowledged by parents as a resource at the service of the transmission of knowledge (see excerpts 20 and 21, and others to be found in Appendix G). The concept of epistemic trust was kept in mind when appraising data, but it was not applied as a thematic code. This apparent absence does not reflect the impression that participating parents were not trusted by their children, but rather reflects a conceptual overlap. As explained in Chapter 3 (section 2.2.2.1) the ToNP provided an important extent of the grounds necessary for the inception of epistemic trust. At the same time, from a theoretical point of view, epistemic trust enables the successful unfolding of natural pedagogy, insofar a child embraces new knowledge signalled by a caregiver's ostensive cues *because* epistemic trust has already been established. In this close relationship, however, it was the ToNP that initially stressed the role of IDSp as an ostensive cue. Accordingly, it was ToNP that was used as a thematic code whenever IDSp was related to the transfer of knowledge— and not epistemic trust, that refers to a more general idea, less delimitated into precise, concrete, and observable behaviours such as IDSp. In other words, epistemic trust as a thematic code was reserved for instances other than ToNP, so whenever parents would mention teaching *without* mentioning IDSp. This almost never happened.

Beyond these technical precisions, the apparently prevalent use of IDSp as an ostensive cue stands amid the parents that participated in this focus group confirms suspicions discussed in chapter 5 (section 2.2). In particular, it downplays the use of IDSp as a source of vocal affective investment while highlighting its role in a different, cognition-related system, focused on the transmission of knowledge. The third study in this dissertation— and its measurement of IDSp —should be valuable in providing further insight regarding the relative importance of affectivity (as in VAW) and cognition (as in ToNP) as parallel, coexisting governing principles that exert an influence on developmental changes concerning parental use of IDSp.

In terms of the unfolding of the focus group in itself, parents such as Paula, Daniel S. and— especially —Dorothy S. tended to capitalize the speaking floor. Although their interventions were informative and often insightful, they left less space for shier peers such as, for instance, Martha or Xavier. Such a disproportion perhaps could have been further avoided through further mediation on the researcher's end. The researcher did not always provide smooth transitions between topics, and sometimes did so without previously discarding that all participants that wanted to contribute to a particular topic had had the opportunity to do so. At the same time, the nature of the event makes it hard to guarantee balanced contributions; while the researcher's questions were studied, their actual wording and timing were spontaneous, as were the participant's answers. Additionally, the fact of hosting a group of busy adults and having one child literally running around, entering and leaving the room made it hard at times for the researcher to focus.

In conclusion, while this first study sheds light regarding our understanding of VAW, it does not suggest any particular connection between a decrease in positive affect and children's interest in music. Duly taking note of these outcomes and considering that positive affect may not be as important as first theorized, I should still proceed to perform a quantitative study that assesses more directly the phenomena which we aim to better understand. Such a study will be reported in the next chapter.

CHAPTER 7. SECOND STUDY: EXPERIENCE SAMPLING OF CHILDREN AGED 18-24 MONTHS

1. Introduction

The first study helped us reformulate our understanding of VAW, emphasizing the progressive expression of negative emotion as a significant change in parental perception of their own prosody as a result of children's linguistic development.

The next discussed step was to assess the everyday musical environment of children 18-24 months-old, either through constant recording— as Vosoughi & Roy (2012) did —, or through everyday sampling methodology such as the one used by Lamont (2008). I will proceed with the second alternative for a number of reasons. First, Vosoughi & Roy's methodology implies major monetary efforts, such as special microphones capable of recording the child's environment all-day-long without suffering the child's manipulation, and paying participants incentives that compensate for such a continuous invasion of their privacy. Additionally, the analysis of all-day-long recordings requires either a machine learning software capable of distinguishing noise from voice, or a large considerable of research assistants that would manually perform such a time-consuming task. Given that none of these resources were effectively available, Lamont's (2008) sampling methodology stood as an accessible alternative.

In addition, experience sampling methodology would allow us to assess occurrences of IDSi, as speculated in the discussion of the first study. Let us recall here that the developmental window of children aged between 18- to 24-months old was chosen because of the changes in parental prosody that the literature suggests take place during this period. In other words, I will take a leap of faith and assume that there is VAW taking place in the lives of children in this developmental window, because the existent literature converges in suggesting so. As discussed in chapter 5 (section 3.5), existing literature suggests that given their age, we can expect participating children to be engaging in musical play and similar forms of synchronous interaction. Nevertheless, following the present dissertation's scope, I will instead search for children's attention to music without (or beyond*)* the appeal of interaction.

The goal of Lamont's original study was to explore "*preschoolers' real life engagement with music in everyday life, examining the choices that they have over music listening and the engagement that they show in relation to music in different contexts*" (2008, p.1). This was carried out by means of a borrowed cell phone that accompanied children throughout the day. Our present goal is similar, yet more specific. Our focus does not lie so much in children's engagement with music in general, but rather in what takes place in the company of their parents. What I would like to quantify is whether at this age— if we assume VAW is taking place —children are already paying sustained attention to music that is *not* embedded in their parents' live IDReg. The reasoning behind this is that because VAW is taking place, children have already deduced that music, unlike language, is a reliable source of conflict-free interaction, therefore boosting their interest in it, even if it is not embedded in the IDReg. For these reasons, certain methodological modifications would need to be implemented. First, because I would like to assess children's behaviour in the vicinity of their parents to check whether involvement of the latter is requested, nothing other than the parent's personal cell phone would be required in terms of devices. Second, because IDSi usually takes place during bedtime, a retrospective question assessing the previous night's bedtime would need to be added. Finally, because I am interested in sustained attention, such a response would also need to be explicitly assessed.

2. Methods

2.1. Sample

2.1.1. Recruitment

Participants (n=25) were previously contacted and invited to participate through e-flyers, posters and leaflets distributed around Cambridge. An academic acquaintance (an established scholar in the field) advised me not to offer any monetary incentive, but instead to make efforts for intrinsically motivating prospective participating parents. A virtual platform was thus created in Facebook and entitled 'Music to my Ears!'. The platform had the following description: "*Music to my Ears is a study in the field of the psychology of music undertaken by Juan-Pablo Robledo, PhD candidate at The Centre for Music and Science, University of Cambridge. This page is a platform for recruiting participants for such a study, and for sharing its results if published".* An e-flyer was posted through this platform in several Facebook groups such as 'Cambridge Mums of Under 5s', 'Cambridge Parents (UK)', 'Mums, dad's., babies and bumps' and further similar ones. The e-flyer invited parents to learn more about their children's musical world. This strategy proved to be the most efficient one as the vast majority of participants were recruited in this manner. Posters and leaflets were distributed in nurseries, primary schools, parish institutions and children's music groups around Cambridge. This strategy turned out not to be very efficient in terms of costs/benefits (see this chapter's discussion) as only one of the 25 participants was recruited by this means.

2.1.2. Demographics

The age of the participating mothers[19] ranged from 32 to 48 years-old with an average of 37.5 (s = 5.2). Their 'target' children (the child in the developmental window of interest) ranged from 18 to 24 months-old, with an average of 20.3 (s = 2.3). Thirteen of them were boys and the remaining 12 were girls, averaging 19.9 (SD = 2.6) and 20.7 (SD = 1.9) months of age, respectively.

2.2. Procedure

2.2.1. Initial questionnaires

Once contacted, mothers were sent an ethical consent from approved by the Faculty of Music. Following their signature, they were asked to complete two questionnaires. The first one concerned basic demographics, as well as six screening questions that assessed their own holistic appraisal of their target child's development (Appendix B). The screening questions consisted of sentences to which the mothers could agree or disagree through Likert scales (e.g. 'My child understand everything I say to her/him' or 'Sometimes my child does not cooperate due to the fact that he/she does not understand what I'm asking for'). The second questionnaire was an online versions of the MacArthur-Bates Communicative Development Inventories (Fenson, Marchman, Thal, Dale, & Reznick, 2007). The MacArthur-Bates Communicative Development Inventories (henceforth 'CDI') is an instrument that assesses of the communicative and linguistic development in young children (Appendix C). It comprises two sections, focusing on lexicon, syntactics and grammatic.

[19]The husbands of two of the total participating mothers occasionally participated in the study, by answering a minor number of phone calls.

2.2.2. Phone calls

Each mother was asked permission to be contacted through her personal mobile phone during one week of her own choice, and only within the ‘sociable hours’ during which they were happy to be contacted. The latter reflected in most cases the total normal hours the children being studied shared with their parents— except bedtime, instance when parents systematically refused to be contacted. Following Lamont (2008), calls were made up to a maximum of three times a day for a period of seven consecutive days at quasi-random intervals, with the restriction that no calls were made less than two hours apart. Over the seven days, attempts were made to sample different times of the day in order to capture the widest possible range of instances. If calls went unanswered, the researcher attempted to call once again no sooner than 30 minutes; a second non-response meant the time sample was discarded and the next call made after a further two hours had elapsed. During each successful call, questions were asked to the adult responsible for the child over the telephone, and data recorded by the researcher. Questions were as follow: 1. Where are you [and child]? 2. What are you and [child] currently doing? 3. How are you two getting along at the moment? 4. Is there any music or singing on at the moment? 5. If not, has there been any music or singing in the last two hours? 6. If so, what is it? (describe as much detail as possible) 7. Who chose it? 8. Is [child] aware of the music? 9. Is [child] responding to/showing an interest in the music? How can you tell? 10. If so, is it only for a few seconds, or is it sustained? How can you tell? 11. Did you play music or sing to your child last night before going to bed? 12. Any other comments?

If no music was currently being played— or had been played —in the last two hours, the interview terminated at question 4. An exception to this was question 11, which was always asked during the first call of the day. The standardized questions would often lead to further ones anytime disambiguation was needed, especially questions 9 and 10, where it was vital for the researcher to guide the mother in distinguishing between sustained attention from mere initial fixation.

2.3. Analysis

As hinted in the fifth chapter (section 3.6), beyond descriptive statistics, the number and heterogenous nature of the intervening variables contemplated in this study justified the use of decision trees. This section will aim to briefly and didactically explain what decision trees— in general — and a Random Forest analysis— in particular —consist of.

Given that the target (dependant) variable is a non-numerical value (a categorical, or qualitative output), and that I wish to predict it by means of the rest of the available variables, the task corresponds to a classification problem (James, Witten, Hastie, & Tibshirani, 2014). In this particular case, the number of variables that may be involved in predicting children's sustained attention is rather high and, as previously discussed, some of them are continuous whereas others are categorical.

2.3.1. Decision trees

Decision Tree Learning uses a decision tree as a predictive model that divides the Predictor Space— that is, the set of possible values for the output variable —into a number of distinct, non-overlapping Regions. Each region is defined by the different variables that predict the output variable. The algorhythm extracts a subset of the total data (subset called 'Training Observations'), generates a number of random trees out of it, and then attempts to predict the behaviour of the output variable in the remaining portion of the dataset. In a classification tree (let us remember that the output variable is categorical), the aim is to predict when an observation will belong to the most commonly occurring class of Training Observations in the region to which it belongs. The goal is also to find the smallest number of regions that reduce as much as possible the Classification Error Rate— the fraction of the training observations in that region that do not belong to the most common class.

Let us put the previous technicalities into concrete terms by means of a hypothetic example concerning the present dataset (Figures 1a and 1b). The model selects a random subset of the 309 phone calls, referred to as Training Observations. The model then randomly selects Child Gender as a variable and divides the Predictor Space into two Regions: Male (R1) and Female (R2). It turns out that out of the total Training Observations where children did pay sustained attention to music other than that sung to (or played at) them live, 32% of the times the child was a girl and the remaining 68% it was a boy, making R2 (gender=male) a strong predictor of sustained attention. The model then randomly chooses Music Genre as a new variable and subdivides the better-predicting Region (R2) into two sub-Regions, depending on whether it was Pop Music (the new R2, 56%) or Nursery Rhymes (R3, 44%) what was being played at the time of the phone call. The total number of Regions then equals 3.

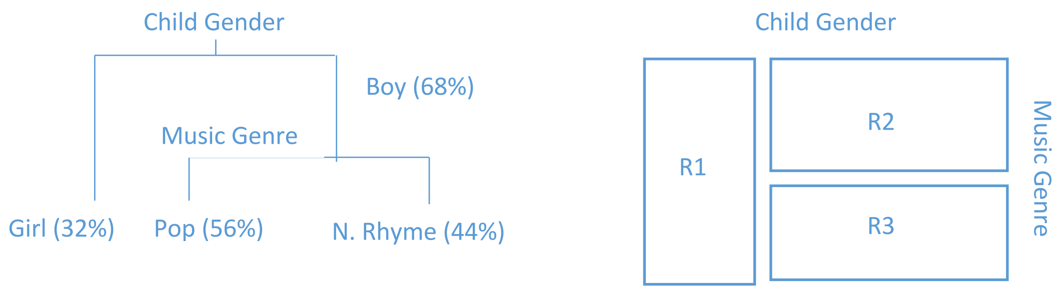


Figure 1a (left). Hypothetical example of a classification tree. The tree has two internal nodes (each bifurcation corresponding to a variable) and three terminal nodes, or leaves. Figure 1b (right). Hypothetical three-region partition of the dataset. R1 corresponds to all phone calls where children did pay sustained attention to music other than that sung (or played) to them live, and the gender of the child was female. R2 corresponds to all phone calls where children did pay sustained attention to Pop music other than that sung (or played) to them live, and the gender of the child was male. Accordingly, R3 corresponds to all phone calls where children did pay sustained attention to a Nursery Rhyme other than that sung (or played) to them live and the gender of the child was male.

It may already be obvious to the reader that in the hypothetic example Music Genre is not as strong a predictor as Child Gender is, thus making this tree less informative that it could be. Indeed, the model will use this tree for predicting the rest of the data set (phone calls that were not used as Training Observations), return a high Classification Error Rate, and thus discard the tree. Several other trees will be generated by the model, with various numbers of nodes[20]. The model will then cross-correlate the totality of the different generated trees, until a final one is chosen for its minimum Classification Error Rate.

Decision trees offer a number of methodological advantages. They are easy to explain and are well suited for handling qualitative predictors/variables. They also allow for simple graphical display, fostering a straightforward interpretation. On the other hand, trees tend not to present the same level of predictive accuracy as some of the other regression and classification approaches. Additionally, trees can be non-robust— a small change in the data can cause a large change in the final estimated tree. Fortunately, by aggregating several independent decision tree models, using methods like random forests, the predictive performance of trees can be substantially improved.

2.3.2. Random forest

The main problem the decision tree model above described presents to current purposes, is that (for reasons that do not concern this explanation) most or all of the trees randomly generated— even if finally discarded —will use the most strong predictor as top node (following our hypothetical example, most trees will use Child Gender as first node), thus overfitting to their training set and eventually constituting a bias. In a Random Forest, each time a node in a tree is considered, a random sample of predictors/variables is chosen as split candidates from the full set of predictors. In other words, in building a random forest, at each node in the tree, the algorithm is not even allowed to consider a majority of the available

[20] The ideal number of nodes in a tree is attained by a procedure called weakest link pruning. As in the case of cross-correlation error, it does not directly concern current purposes, so its principles will not be explained.

predictors, consequently decorrelating the possible trees. In the hypothetical example, a Random Forest would only allow a reduced number of randomly generated trees to use Child Genre as predictor, so that other predictors will have more of a chance.

Once the totality of the forest (the ensemble of trees) has been correlated, a comparison of the importance of a given variable in each tree can be performed. The Mean Decrease in Gini represents such a comparison, and consists of the average of a given variable's total decrease in node impurity, weighted by the proportion of samples reaching that node in each individual decision tree of the random forest. This is effectively a measure of how important a variable is for estimating the value of the target variable across all of the trees that make up the forest. In simple words, a higher Mean Decrease in Gini indicates higher variable importance.

2.4. Hypotheses

1) Children of this age would, even in the company of their parents, pay autonomous, sustained attention to music outside de IDReg. This scenario would imply an intrinsic motivation unmediated by their parent's engagement.

2) Alternatively, children of this age would pay sustained attention to music outside de IDReg, mainly if their parents are also engaging with the music. In this scenario, there would be an interest in that is not sung by their parents, but such an interest would be— indirectly —sustained by the latter's engagement.

3) As a third scenario, children of this age would show no motivation for listening to music that is not yet embedded in the IDReg.

3. Results

3.1. Music

Results were obtained from all 25 mother-child dyads, as no one offered less than 10 data points. This provided a total of 431 recorded episodes, with an average of 15 per dyad (SD = 2.2). Of the total 431 episodes sampled, a majority (69%) occurred at home, followed by 12% taking place in a vehicle— usually in the car, but sometimes walking (Figure 2). Other, less-common locations included public spaces such as libraries or shops, or parks and clubs (1.9%).

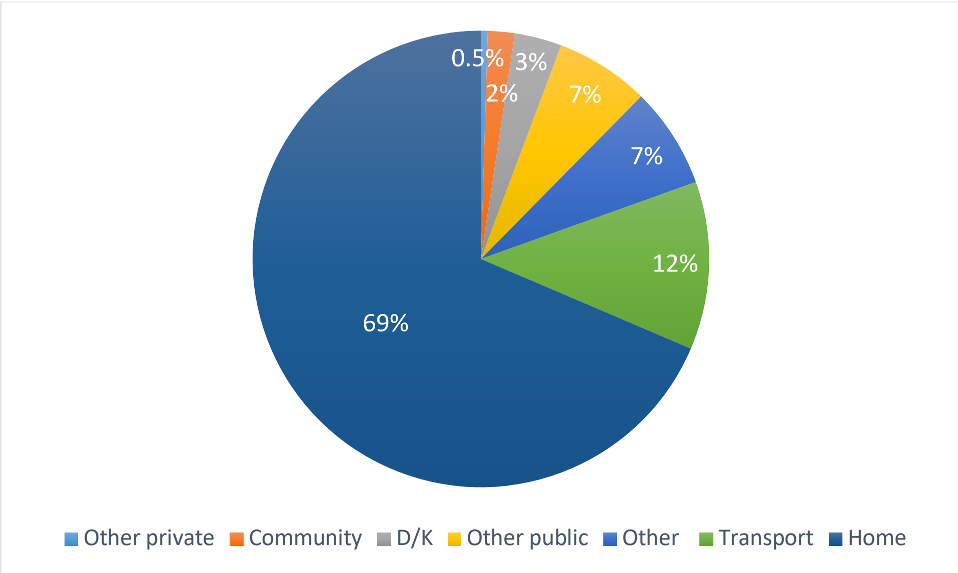


Figure 2. Location of the child during phone calls. 'Community' stands for public spaces where participants often meet members of their community, such as such as parks and clubs. 'Other public' refers mostly to libraries or shops, 'Other private' refers to friends' or relatives' houses. Finally, 'D/K' refers to episodes where the location could not be defined.

Out of the total 431 episodes, 355 of them (82.4%) had some form of music taking place during them (Table 1). Amid these 355, 141 (41%) corresponded to episodes taking place during the call and 214 (59%) to situations that had taken place during the preceding two hours. Music was chosen mainly by the mother or the child itself, with rather even proportions of 36% and 35% respectively (Figure 3). Siblings (11%) and fathers (6%) did so less often, as well as friends, relatives, or radio programmes (12%).

	#	%
Episodes with music	355	82.4
Episodes without music	76	17.6
Total	431	

Table 1. Episodes with or without music taking place, and corresponding percentage out of the total.

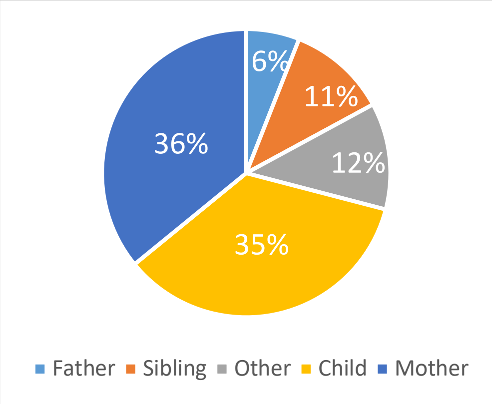


Figure 3. Displays who chose the music taking place during the episodes. 'Other' stands for people other than the ones mentioned as well as instances when it was the radio's programme.

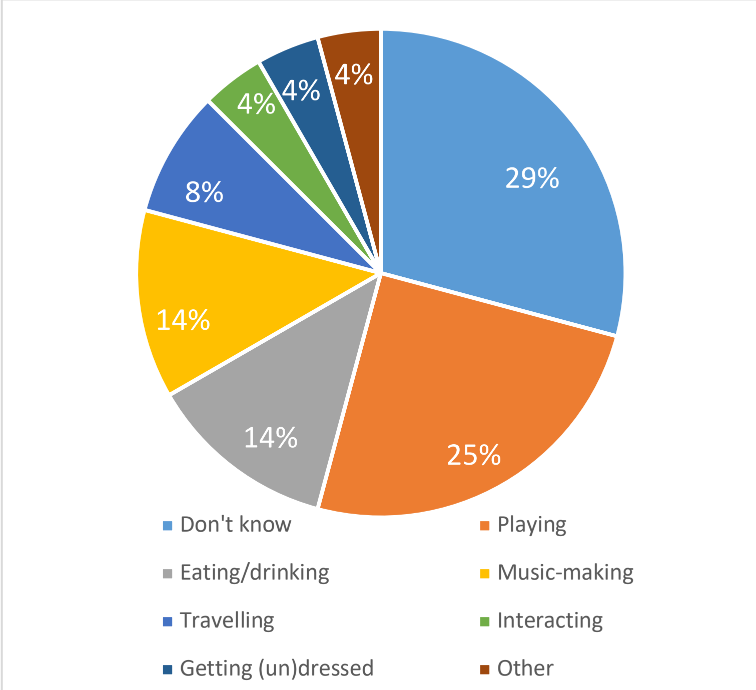


Figure 4. Activity taking place right before the child started self-singing.

Forty-six of the total episodes containing music (12.9%) consisted of music spontaneously sung by children themselves. This excludes instances when the child would sing as a reaction to other forms of music taking place at the time. Self-singing took place almost invariably at home (94% of the episodes that contained it). Self-singing occurred most often while playing (25%), followed by eating or drinking (14%) and music-making (14%) (Figure 4).

As previously mentioned, mothers were asked once a day whether there had been singing at bedtime the night before. Singing at bedtime had taken place 56% of the assessed nights.

3.2. Attention to music

Nearer to the main research question of this work, the episodes of most interest are those in which mothers reported to have the impression that their children were paying sustained attention to recorded music. Let it be reminded, sustained attention is whenever a stimulus prompts an attentional bout longer than an initial fixation. Initial fixation is a short, few-seconds-long attention bout that is directed to any novel stimulus in order to assess it, not necessarily meaning that the stimulus is of any interest. Let it also be clarified, self-singing does not count as sustained attention to music. During phone calls, whenever parents would report sustained attention, it was asked from them to point out exactly what behaviour they relied on to generate such a judgement. Parent's interpreted a number of behaviours as signs of sustained attention. The main one was staring at the source of music. Other prevalent ones were body engagement (dancing, swaying, rocking, doing the song's actions, etc.), singing along, or suddenly interrupting the task the child was involved in when becoming aware of the music.

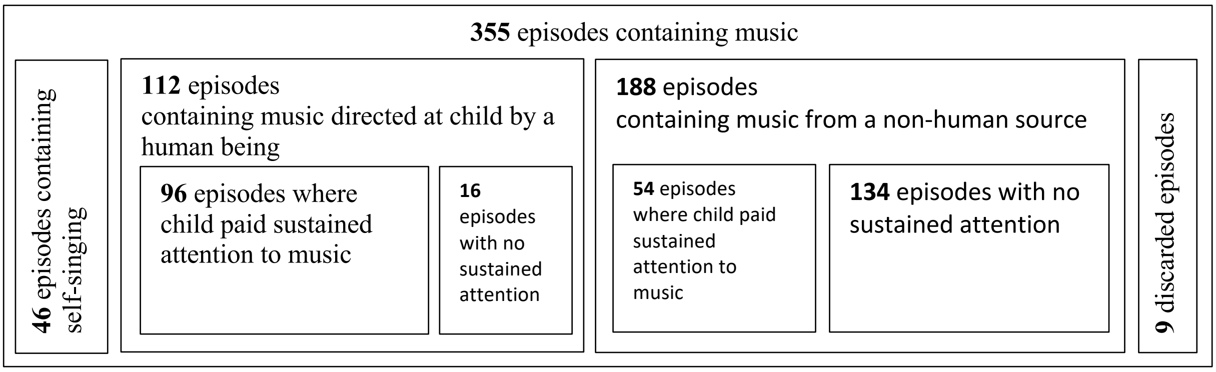


Figure 5. Schematic display of the distribution of episodes regarding the source of music presented to children (human vs non-human) and whether such music prompted or not the child's sustained attention. The size of the boxes is not accurately proportional, but approximate and referential.

Out of the 309 episodes that contained music other than the child's own singing (355 minus 46 episodes of self-singing), 112 episodes corresponded to music sung (or in very few cases, played) to them live by a human being, and 188 episodes contained recorded music

(See Figure 5). Within the latter, only in 54 episodes did the mothers report their children paying sustained attention to music (28%), while most of the times such music prompted only initial fixation or no visible interest at all (72%). A contrary proportion can be acknowledged in the case of music performed live to children by a human being, which prompted the children's sustained attention with little exception— 96 episodes, 85.7%.

Figure 6 displays the different activities taking place right before the moment children paid attention to music that was performed live for them, as well as their relative proportion in terms of percentage. Interacting with an adult (usually one of the parents, most often the mother) was the most prevalent scenario (27%), closely followed by playing (25%). Eating and drinking, bathroom activities, and travelling also presented some minor prevalence.

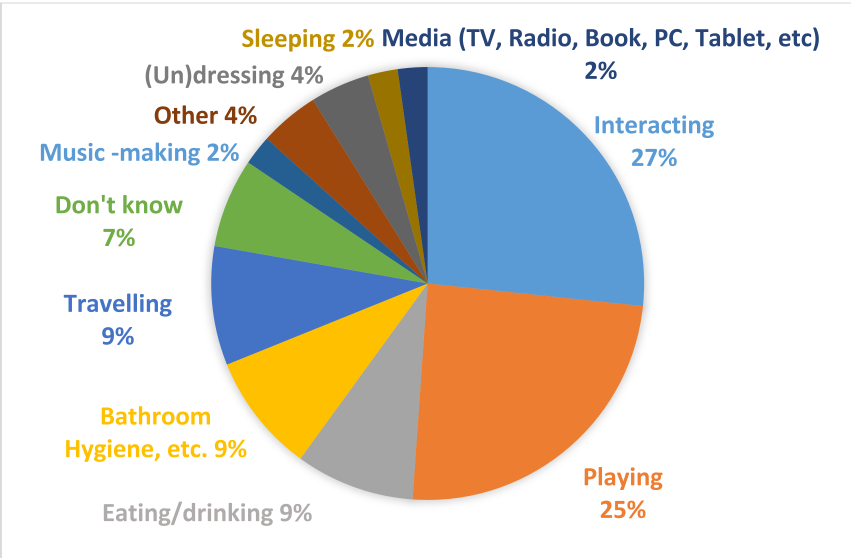


Figure 6. Activity taking place right before the moment child paid attention to music that was performed live by a human being

Mirroring the contents presented in the previous paragraph, Figure 7 displays the different activities taking place right before the moment children paid attention to music that— in this case— was *not* performed live for them but played by a non-human source. Their relative proportion in terms of percentage is also shown. Playing was the most prevalent scenario (30%), followed by instances when the child was engaged with some form of audio-visual media that comprised music as part of it such as tablets, computers and children's books (22%). Travelling (15%), eating or drinking (7%) and bathroom activities were also mentioned (4%).

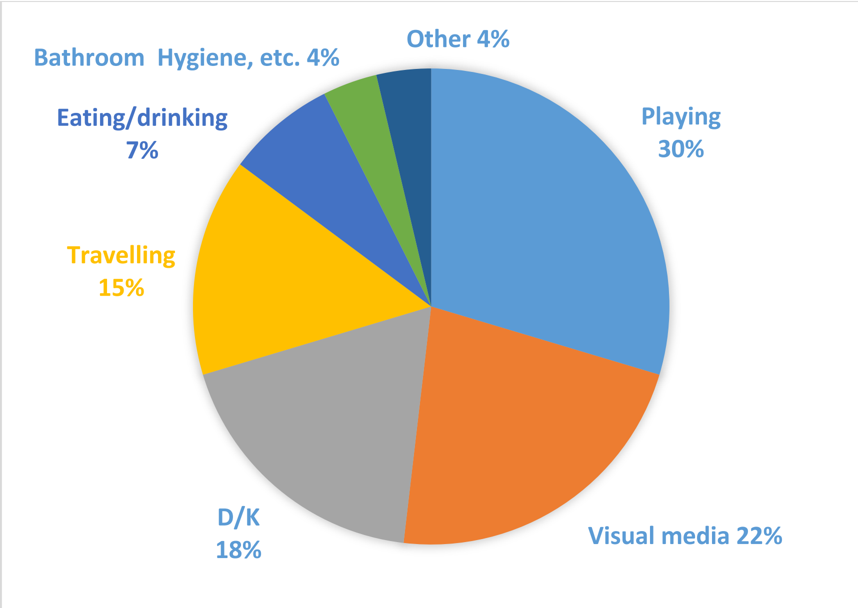
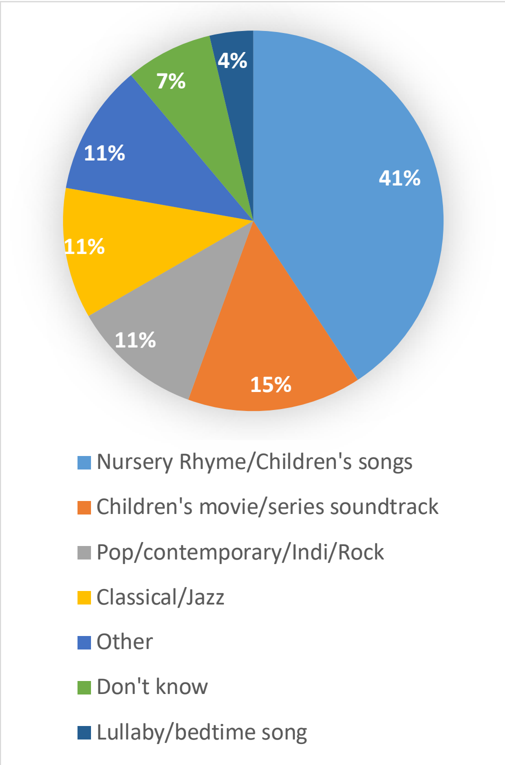


Figure 7. Activity of the child while paying sustained attention to music that was not being directed to them live by a human being. D/K stands for episodes when the mother could not specify an activity.

If home was the most prevalent location for episodes during which there was music playing and for episodes in which children paid sustained attention to music performed live, it is even more the case for episodes in which children paid sustained attention to music played by a non-human source. As seen in Figure 9, the most common location was home (78%), followed distantly by a means of transport —almost invariably a car.

Finally, Figure 8 displays the genre of the music to which children paid sustained attention to (and was not being performed live by a human being). Almost half of the instances (41%) correspond to nursery rhymes and children's songs alone. Soundtrack music— part of an audio-visual stimulus —comes next with 15%. Contemporary music (indi, rock, and mostly pop music) follow with 11%, as do classical music and jazz.



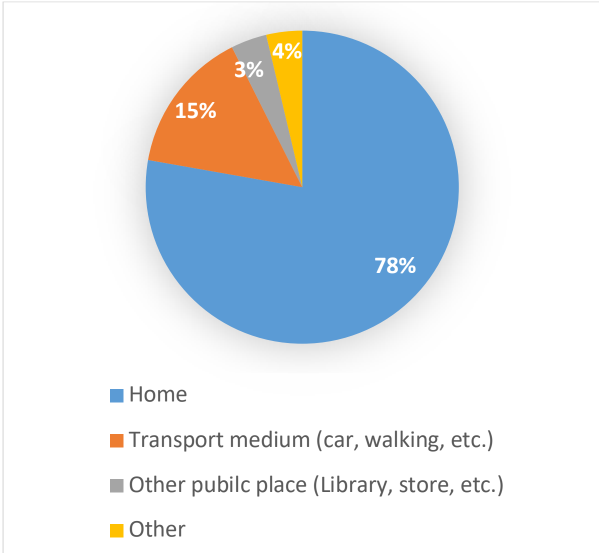


Figure 8 (right-hand side). Genre of the music to which children paid sustained attention to (and was not being performed live by a human being).
Figure 9 (above). Location of the child while paying sustained attention to music that was not performed live by a human being).

3.3. Conflict

Signs of mother-infant conflict were extracted from answers to phone question #3 (*How are you two getting along at the moment?*) as well as spontaneous interaction during the phone call. Conflict— if any —was classified in two respects: the source of conflict and the mother's reaction to it. Sources of conflict were assigned into one of the three most common types : child demands for attention (e.g. crying, fretting or clinging), protests or disobedience, and sleeping problems (Ullstadius, 2014). Mother's reactions to conflict were also assigned to one of three categories (Rijt-Plooij & Plooij, 1993). The first one is annoyance, where the mother is confronted with the child's regressive behaviour but tries not to express such feelings. In this category, the consequence of the conflict in interests remains hidden from the infant and locked up inside the mother. A second category is promoting progress in independence, consisting of episodes where the mother did not simply show signs of being annoyed, but instead ignored and/or rejected the annoying behaviour. Promoting progress in independence also includes episodes where the mother acted upon the situation by means of different strategies, like distracting the infant or teaching them new skill(s) or practices that could replace the annoying behaviour. The third category corresponds to mother-infant clashes, which encompasses two cases. First, whenever the infant resisted maternal attempts to promote progress in independence by mean of screaming/yelling, trying to generate physical damage, and/or throwing a temper tantrum. Mother-infant clashes also accounts for episodes where the infant appears to wish to increase maternal attention and/or proximity at a moment when the mother is busy, by taking the initiative in a demanding/quarrelsome way, as if 'out of the blue'.

As seen on Tables 2 and 3, 44 of the 432 assessed showed signs of some sort of conflict— around 1 every 10 episodes. The most common conflict source was demands for attention (6.5%), followed by protests and disobedience. There was virtually no conflict due to sleeping problems. In terms of the mother's reaction, promoting progress in independence was the most prevalent outcome (5.1%), followed by annoyance (3.3%) and mother-infant clashes (1.6%). No correlation was found between conflict and the ages or the linguistic development of the children assessed.

Source of conflict	Occurrences	%
Demands for attention	28	6.5
Protests or disobedience	14	3.3
Sleeping problems	1	0.3
Total	44	10

Table 2. Episodes containing conflict, classified in terms of the latter's source. The '%' column displays what percentage out of the total 431 episodes the figure in the 'Occurrences' column represent.

Mother's reaction	Occurrences	%
Annoyance	14	3.3
Promoting progress in independence	22	5.1
Mother-infant clashes	7	1.6

Table 3. Episodes containing conflict, classified in terms of the mother's reaction to it. The '%' column should be interpreted same as in Table 10.

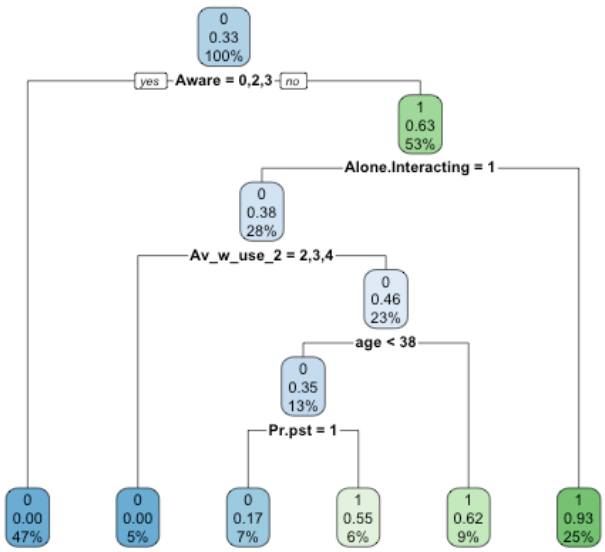
3.3. Factors explaining sustained attention to music. Random Forest analysis

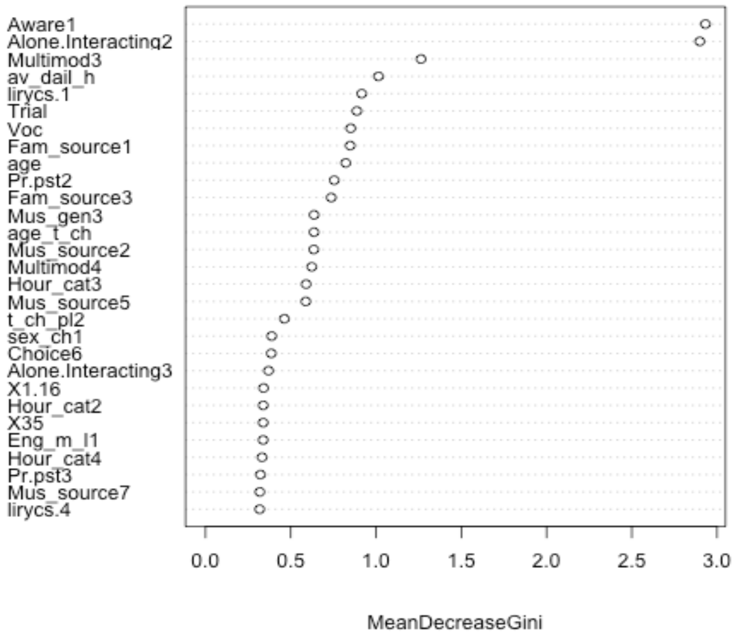
Having separately reported the different elements that concerned the present study— language acquisition, conflict, and sustained attention to music —, it is now possible to assess their potential interrelation. In other words, having described the available data, it is possible to assess which variables— if any —may predict children' sustained attention to music other than that sung (or played) to them live (Appendix D). Some of them are continuous (e.g. children's age) whereas some are categorical. Because during phone calls parents could not report exactly how many seconds of attention did their children pay to a given piece of music, such impression had to be considered categorically: either they paid sustained attention to it, or they did not.

Random forest results

A Random Forest analysis was run using default parameters as specified by the R RandomForest Package with 500 trees and a random sample split parameter of sqrt(p) (where p is the number of variables in the dataset). The decision tree seen on Figure 10 represents just one of the many in the total generated forest. It shows that, within the dataset, a child paying sustained attention to music can be predicted mainly through five variables (the internal nodes in the tree): whether the child was aware of the music; whether the context of in which the child paid sustained attention to the music involved interaction with somebody, the child's linguistic development in terms of grammar, the age of the mother, and whether the parent was reporting music being played at the time or during the last two hours.

However, the critical result of the random forest is what can be deducted from the sum of all 500 trees, such as the variables that most trees used as nodes, and the purity of the latter. Such information is presented through a variable importance plot (Figure 11), in which the most relevant variables are displayed according to their Mean Decrease Gini (see previous section).





Figures 10 and 11. Figure 10. Example of a decision tree. Figure 11. The variable importance plot shows how important each variable is when classifying the data. The predictor variables are on the y-axis, with the Mean Decrease Gini on the x-axis. The Mean Decrease Gini is a measure of how each variable contributes to the purity on each node in a tree. The first fours variables are hereby listed. Aware1 = The child is aware of the music. Alone.Interacting2 = The child was interacting with someone when the phone call took place. Multimod3 = The music is rendered in a multimodal context, engaging all of the child's senses. Av_dail_h = Average daily hours the mother spent with the child. Because only the first two variables are relevant and worth explaining in this legend, the rest of the variables and their explanation can be found in Appendix D.

The variable importance plot shows that only two variables detach from the rest: whether the child was aware of the music, and whether the child was alone or interacting. More precisely, it shows that the child being aware of the music is the best predictor. Although the latter is obvious, it does give an idea of the biggest possible Mean Decrease Gini in this random forest. The plot also shows that the child being in a context that involves interaction with another person is almost as strong a predictor of sustained attention. The rest of the variables did not present a clear predictive power.

4. Discussion

The most relevant finding in terms of the main thesis is that data allows to confirm the second hypotheses laid out in the introduction while discarding the other two. It was hypothesized that children in this developmental window would pay sustained attention to music outside de IDReg, mainly if their parents are also engaging with such music. The variable importance plot (Figure 11) exhibits precisely such an outcome: the only variable that reliably detaches from the rest in predicting sustained attention in our sample is interaction with a significant other— mainly, the mother. This strongly suggests that although there are clear tendencies in terms of music genre (nursery rhymes), location (home) and activity (playtime), none of these specificities are strong-enough, on their own, for systematically prompting sustained attention to recorded music.

The Interaction factor also evidenced far more predictive power than the Trial and Present-Past factors, outcomes that carry methodological implications. Let us first examine the Trial factor results, and its potential effect on learning curves and social desirability. The Trial variable reflects the temporal location of the phone call within the total seven days that the study lasted for. For each participant, the first successfully-held phone call corresponded to trial n.1, the second to trial n.2 and so on and so forth, until potentially reaching the twenty-one trials limit (three phone calls a day for seven days). Because it was the mother who was asked to judge and report her child's attentional state, a learning curve or 'carry-over effect' could be expected: the more trials took place, the better they understood what was asked from them. Furthermore, a social desirability effect could have led them to give the researcher whatever they thought he wanted from them. Thus, it remained a possibility that reported episodes of sustained attention could be mainly explained by the amount of trials mothers had participated in— the larger the trial number, the bigger the possibility of reporting sustained attention. In such a case, however, the Trial factor would have detached from the rest of the factors in the variable importance plot, and this was not the case. As a result, although a certain learning curve and social desirability effect can indeed be acknowledged, results do not support their exerting a significant influence on the mother's judgment.

A similar outcome was found when examining the influence of the Present-Past factor. As detailed in the Procedure section, parents were asked in each phone call whether any music or singing was taking place at the time. If it was the case, then the number 1 was assigned to the trial, indicating that music was being played presently. On the other hand, if music was not being played currently but had been in the last two hours, the number 2 was assigned to the trial, indicating that music had been played in the past. The crucial difference between these two kinds of reports is that in the latter, mothers relied on their memory, their recollection being perhaps more prone to the pressures of social desirability. It therefore remained a possibility that reported episodes of sustained attention mainly corresponded to events that had taken place before the phone call. Once again, however, the random forest analysis discarded such a possibility. As seen in Figure 11, the modality 'Pr.pst2' (trials reporting music that had been played in the past) did— as expected —have a certain impact on the probability of sustained attention being reported. Nevertheless, such an impact is not significantly larger than that generated by the vast majority of the accompanying factors. This result does not support the idea that the exercise of recalling made a significant impact on mother's reports, nor does it justify a need to analyse data separately in terms of the Present/Past factor.

Another relevant outcome in terms of discussion is that we are in a position to, at least partially, compare the locations, activities and other specificities of this sample of children aged 18-24 months-old, to Lamont's (2008) findings concerning children 3.2–3.9 years-old. It is important to keep in mind that in the 2008 study, 19% of the episodes happened at nursery, while in the present study such a venue was never considered. Nevertheless, In Lamont's study, nursery reflected a higher proportion of the sample's no-music episodes, so it can be well argued that the main location to focus on in terms of analysis is the children's homes. In the 2008 study, out of a total of 437 episodes, 81% had music exposure. Given that the present findings are drawn from 355 episodes containing music— 82.4% of the 431 total episodes—, adding both studies' data suggests that children's exposure to music (regardless of the music's specificities such as genre, location, etc.) does not change significantly from one examined developmental window to the other. In other words, both children 18-24 months-old and 3.2–3.9 years-old are exposed to music constantly— up to four fifths of the time.

As Lamont reports, the main activity which accompanied the 353 music episodes was 'entertainment', occurring 30% of the time and mainly consisting of watching television or videos. In the present dataset, when looking at the activity of the child while paying sustained attention to music that was *not* being directed at them live by a human being (Figure 7), visual media (22%) is not the main activity, but general play (30%). The latter suggests that children 18-24 months-old are not yet as familiarized with music stemming from television and videos as 3.2–3.9 years-olds are.

Perhaps the most interesting contrast between the two populations concerns the parent's own musical performance. As discussed in chapter 5 section 4.2, Lamont (2008) reports a scarcity of live music performed to children by others, with only three live music episodes involving a father singing at bedtime, a street busking group, and a sibling playing the piano. Something similar is reported in a longitudinal study involving 10 families, during which live singing was found to be much rarer than using music to accompany an established activity (Custodero, 2006). A rather different scenario is portrayed by the present dataset. Out of the 355 episodes containing music, 122 of them episodes contained music directed at the child by a human being, constituting a 34.4%. Thus, while children 3.2–3.9 years-old rarely obtain music from a human source during daytime, their 18-24 months-old counterparts are still receiving live performances around one third of the time they are exposed to music as a whole during daytime. To such an amount of live music, instances of singing at bedtime should be added, the latter taking place 56% of the assessed nights. Bedtime singing could have perfectly occurred amid the population assessed by Lamont (2008); however, such instances were not systematically contemplated due to the author's decision to focus exclusively on 'sociable hours'. It therefore remains an open question to what extent 3.2–3.9 years-olds receive live music before sleeping.

Assessments of conflict returned no evident relationship to sustained attention, as reflected both by the descriptive statistics and the random forest analysis. Overall, there was very little of it, and it was always well-handed by the participating mothers. The most common conflict source was demands for attention, followed by protests and disobedience, which seems like a natural outcome considering the study's design, in which parents attention was capitalized by the phone calls. In terms of the mother's reaction, promoting progress in independence was the most prevalent outcome. Results, both in terms of conflict source and mother's reaction, could be hypothesized to— at least partially —be the result of social desirability. Indeed, whenever conflict took place, mothers were on the phone with the researcher, the latter inevitably representing a potential source of moral judgement.

As a whole, information presented in this discussion do allow us to step forward in our understanding of children's motivation for paying sustained attention to recorded music. Current data suggests that children 18-24 months-old are still receiving plenty of music directed live at them by a significant other. Perhaps because of the same reason, music that is *not* directed live at them is not yet worth their sustained attention, unless such music takes part of a triadic situation in which a significant other's attention is also involved. In both cases, it seems to be the adult's attention directly (in the case of the parent singing) or indirectly (when jointly listening) pointed at the child that makes recorded music relevant. This could be interpreted as music serving as a source of positive affect, concretely experienced in company of the parents. In other words, the transfer effect between IDReg and recorded music (Chapter 5, section 3.2) is still concretely mediated by the parent's presence. As a consequence, children have not yet internalized this relationship and therefore need to be positively reinforced through music.

The caregiver's attention directly or indirectly pointed at the child can make recorded music accessible, rather than relevant. In this sense, findings above can be interpreted in terms of attachment and parent-offspring conflict, epistemic trust, and cultural learning (Tomasello, Kruger, & Ratner, 1993). In terms of attachment, joint listening constitutes a form of interaction that, like singing, conveys proximity and attention, therefore in turn generating perceived safety. In terms of parent-offspring conflict, in this developmental stage recorded music allows for a decrease in parental investment. The larger investment that IDSi represents as an active task that requires fine-grained vocal and cognitive effort seems to be at least partially replaced by the comparatively-less costly behaviour that jointly listening to an external source of music represents. By this token, parental investment is successfully diminished while at the same time avoiding confrontation with the child.

But perceived safety is not the only function of attachment. As discussed in chapter 3 (section 1.3), a secure bond also provides the emotional base that allows a progressive emergence in the child of an autonomous motivation for the exploration of the surrounding world— including music. In this context, mentalisation (a caregiver's capacity for acknowledging their child as having an internal, mental life that includes desires, and understands their behaviour as being driven by such) illustrates one way in which motivation stems from attachment. As the child's mental states are recognized and reflected by means of the caregiver's efforts, the child itself gains recognition of such states. It seems probable that children at this age are still not entirely familiar with culturally-appropriate responses to music (music still presents itself to them as partially opaque), and perhaps with their own spontaneous responses, emotional or otherwise. Facing such challenges, it might be the case these children are still in need of their caregivers' mentalisation. At the same time, since music and related expected behaviours might still be opaque (not obvious or self-explanatory), their resorting to the caregiver as a source of knowledge stands as an instance of epistemic trust.

Finally, let us examine results in terms of cultural learning. Tomasello and collaborators conceived the phenomenon as a capacity for acquiring culture. In cultural learning, children do not simply focus their attention on the location or object of another individual's activity, but rather attempt to appraise the situation from the other's perspective. In this case, recorded music can be considered the object of attention, and it could be argued that the child pays attention to it in an attempt to understand the adult's interest in engaging with it, what their motivation for listening to recorded, non-live music might consist of, as well how does that motivation manifests in terms of behaviour. Findings could thus be interpreted as a form of cultural learning in particular and, therefore, part of the Western enculturation process in general. However, as compelling as this idea might be, the present study provides no direct evidence for it. Given that cultural learning works on the basis of joint attention (Tomasello, 1988), an assessment of children's gaze would need to take place in order to confirm that it is cultural learning that is taking place during interaction with a significant other, an assessment that should be included in this thesis' next study.

Once again, it seems to be direct or indirect parental attention that prompts children's sustained attention to music, and not any kind of direct or indirect relationship to conflictive interaction as described by Ulstadius (2014) or Rijt-Plooij' and Plooij (1993). Such outcome does not point in the direction of any direct relationship between VAW and children's motivation for paying sustained attention to recorded music. In theory, VAW comprises an increase in parental vocal expression of negative emotions. Because I assumed— but did not measure —any parental prosodic data, following the main thesis we could at least have expected some degree of correlation between conflictive behaviour and the children's sustained attention to recorded music. In other words, we could have expected that in mother-infant dyads where more conflict could be acknowledged, the child's motivation for autonomously listening to recorded music would have been accordingly greater.

Now that the present study's findings have been properly acknowledged and discussed, I are in conditions of further refining the assessment of our thesis. First, because a) data suggests that children 18-24 months-old are still receiving plenty of music directed live at them by a significant other, b) the transfer effect between IDReg and recorded music is still concretely mediated by the parent's presence and therefore c) at this age children are not yet motivated to paying sustained attention to music that is *not* directed live at them, it seems like our inquiry should shift towards children older than those of the current sample. Second, because some of the specificities concerning the actual location, activity and music genre that characterized music listening in this study resonate with Lamont's (2008) findings, our next piece of research should focus on music listening that takes place at home, while playing, and involving nursery rhymes. Last but not least, the next study should focus solely on music that is *not* directed live at children. Indeed, because I are interested in the onset of children's motivation for paying sustained attention to recorded music, a methodology should be devised where parental singing is avoided while children's autonomous engagement with recorded music is privileged.

A brief note on the hardship of participant recruitment should be added. As mentioned in the section 2.1.1 of the present chapter, posters and leaflets were distributed in nurseries, primary schools, parish institutions and children’s music groups around Cambridge, proving to be costly in terms of effort and— especially —time, while translating into virtually no participants. The combination of being foreign and male seemed to be a disadvantage. In our society, people still tend to associate childcare with the female gender— let us remember that on one exception, all participants were mothers. At the same time, people tend to trust more easily somebody that shares their own culture and language. This can be accentuated when it comes to exposing their (or others’) offspring to potential danger. After all, it is true that the researcher got only progressively acquainted with the verbal and nonverbal cues that the British use for conveying trustworthiness. Thus, while people in the mentioned institutions (e.g. the first person to open the door, the person in charge, and potential participants) welcomed the researcher politely, they were nonetheless confronted with a rather rare (and potentially hazardous) situation. In the light of these experiences, participant recruitment in further studies should privilege virtual alternatives.

CHAPTER 8. THIRD STUDY: SIX LONGITUDINAL CASE STUDIES OF CHILDREN AGED 15- AND 23-MONTHS-OLD

1. Introduction

The following report summarizes the main findings of this dissertation's third and last study. Results from the second study (which consisted of an experience sampling three times a day for a week involving 25 participants) showed that active engagement with an adult—almost invariably the mother —was the main factor explaining 18-24 months-old children's sustained attention to recorded music. Results showed no connection between children's sustained attention and their level of linguistic competence— the latter devised as a proxy for the parent's use of the IDReg.

There were three main changes to be applied in a following study. First, it should focus on children roughly the same age but also slightly older than those in the second study's sample. Second, it should focus on music listening that takes place at home, while playing, and involving nursery rhymes. Third, the next study should focus solely on music that is *not* directed live at children— our main interest. To such changes, three critiques could be added. First, the second study's data came from a single week, impeding the assessment of any within-subject development. Second, it provided only an initial assessment of the children's general linguistic development, without focusing on pragmatics and its temporal evolution. Finally, it did not actually assess the parent's speech, infant-directed or not, but instead largely assumed there were changes occurring in that developmental window and sought to assess them by means of a proxy. A third, longitudinal study was therefore devised, that would directly assess children's sustained attention to music, pragmatic competence, and parental prosody. The study would focus on a handful of cases, so as to allow the addition of qualitative observation.

It must be bore in mind that a setting was needed where children could spontaneously— but would not be externally prompted to —engage with recorded music. This was a subtle task, as listening to music is often accompanied by many other forms of music-related dynamics and activities. The key lied in the temporal *sequence* of events. If music is presented *along* with an invitation to supported singing, dancing, playing basic instruments, or other activities described in previous chapters, it will be impossible to discern whether it was music itself (the auditory content) or social interaction that motivated the child's attention. On the other hand, if music is casually presented without further stimuli, the child may or may not pay sustained attention to it and may or may not spontaneously wish to *eventually* engage in additional musical activities to its sound. In this sense, children's sustained attention to music *may* be supplemented by musical play. Let us remember, Marsh and Young (2016) defined musical play as activities that are self- initiated, enjoyable and intrinsically motivated— led and controlled by children themselves (chapter 5, section 3.5). Following this logic, singing, dancing, playing instruments or any other parent-child musical play activities may perfectly take place during the study, as long as it is children that resort to them as their spontaneous response to recorded music. In order to provide such a setting, this third study will introduce a novel methodology: a one-way remote video recording technique (ORViRT).

The literature indeed suggests that social interaction is a key factor behind children's interest in music (e.g. Mehr 2016, see Chapter 5, section 4.1). As explained in this introduction's first paragraph, such a notion was largely confirmed by the second study. Therefore, the present study will introduce a further attempt to isolate and gauge the relative importance of social interaction over surrounding factors— notably, changes in parental use of IDSp.

2. Methods

2.1. Sample

2.1.1. Recruitment

Following recruitment experiences detailed in the previous chapter's discussion, the present study was circulated solely by means of e-flyers. Following advise received when advertising the second study, this time the e-flyer invited parents to join a 'musical adventure' along with their children. Over forty parish institutions were contacted, with a handful of them actually distributing the material through their mailing lists. Such a strategy did not return any participants whatsoever. The e-flyer was also posted via the Music to my Ears! platform in the Facebook groups mentioned in the previous chapter. One mother-infant dyad was recruited in this manner. Finally, the e-flyer was distributed at a broader level in the UK by means of paid Facebook adds. Several mothers expressed mistrust of the researcher's intentions, and posted comments warning their fellows not to participate. Accordingly, the original e-flyer had to be amended and re-posted in order to make it explicit that at no point during the study children would be left alone in the virtual presence of the researcher. Five out of the six mothers that finally participated in the present study were contacted by means of these adds.

2.1.1. Demographics

Six mother-infant[21] dyads were recruited. Three of them included a child around 15-months-old, and three of them included a child around 23-months-old at the study's start point. All mothers and their children were native English speakers. Participants were offered £100 in compensation for a 20-week commitment. In the case of two participating dyads belonging to the 23-months-old group (case studies 5 and 6), the study lasted for 10 weeks.

[21] Advertisements invited parents regardless of sex and gender. However, as in the case of the previous two studies, only women showed interest in participating.

This decision was made after around a month of assessment, during which the researcher had gathered the impression that the engagement of these children (the oldest in the study) with recorded music was markedly and consistently different from the rest.

2.2. Variables

Tables 1a and 1b show the list of variables involved in the study's design, as well as the instruments employed for their assessment and the frequency of the latter. Assessment instruments will be described in detail in the next section.

The literature in chapter 5, section 4.3.2 predicts that children's linguistic competence should prompt quantal changes in parental prosody. Such changes have been considered to be part of VAW. Accordingly, as seen on Table 1a, children's pragmatic competence was assessed on fortnightly basis by means of the Language Use Inventory (LUI). Depending on this first variable is parental prosody, assessed weekly through the recording of an everyday routine in the life of the mother-child dyads. The main thesis predicted that changes in parental prosody (VAW) would in turn motivate children's engagement with recorded music. Accordingly, as seen on Table 1b, maternal prosody stands to this purpose as the independent variable, with the child's attention to music depending on it. This last variable was assessed weekly by means of a one-way remote video recording technique (ORViRT).

a)

Variable	Type of variable	Instrument	Frequency of assessment
Pragmatic competence	IV	LUI	Fortnightly
Parental prosody	DV	5-min routine audio recording	Weekly

b)

Variable	Type of variable	Instrument	Frequency of assessment
Parental prosody	IV	5-min routine audio recording	Weekly
Attentional bout duration	DV	ORViRT	Weekly

Tables 1a and 1b. Variables in the study's design. The acronym 'LUI' stands for Language Use Inventory, and 'ORViRT' for one-way remote video recording technique.

2.3. Assessment instruments

2.3.1. Initial questionnaire

During the first week of the study, participating mothers were asked to complete a custom-made online initial questionnaire through an internet link generated by the AllCounted survey system (https://www.allcounted.com). The questionnaire assessed participants' demographics, information about the participating children's family composition (Table 2), as well as mothers' impressions concerning their children's development (see each individual case, in section 3.1 in this chapter). The complete questionnaire can be found in Appendix E.

	P1	P2	P3	P4	P5	P6	Average
Age	31	27	27	33	36	35	31.5
Latest qualification[22]	7	3	3	7	6	8	5.6
Occupation hours per week	1	4	1	0	9	8	3.8
Average daily hours spent with the target child	14	12	12	10	4	4	9.3
Total number of children	3	1	1	2	1	1	1.5
Target Child's ordinal position amid siblings	3	1	1	1	1	1	1.3

Table 2. Participants' demographics.

2.3.2. Language Use Inventory (LUI)

The Language Use Inventory for Young Children (LUI) is a standardized parent-report measure that assesses a child's pragmatic language development or, in simpler words, how children use their language in everyday settings in interaction with other people (O'Neill, 2007).

The LUI consists of 14 subscales divided into three parts (see Appendix F). Part 1 assesses children's communication with gestures. Part 2, the child's verbal communication concerning not just vocabulary itself but also what is actually achieved through words. This part assesses, for instance, how children use words to get people to notice something, and

[22] Figures in the Latest Qualification row correspond to those stipulated by the British government (https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels) and described in detail in the Demographics section of this dissertation's first study.

about the child's questions and comments about themselves or others. The last part of the questionnaire assesses children's use of longer sentences. The questionnaire returns a total raw score. Whenever children were 18-months-old or older, scores were transformed into an age percentile. Participating mothers completed an online version of the LUI every two weeks. In some occasions, mothers would forget to complete it, assessments therefore being three (instead of two) weeks apart. The percentage of questionnaires completed in schedule was 82%.

2.3.3. Routine audio recording

2.3.3.1. Pre-processing and labelling

Participating mothers sent through their mobiles phones a weekly 5-min audio recording of the same routine. The mothers of case studies 5 and 6 proved to be particularly busy (one was a single mother and the other a researcher), and sometimes spaced their recordings more than one week apart in spite of the researcher's reminders. Routines consisted of the child playing, eating, getting dressed or changed, having a bath, and the occasional joint reading.

The average duration of such recordings corresponded to 5'10'' (SD = 36''). By means of the software Praat (Boersma & Weenink, 2019), both the child and the mother's utterances were extracted and labelled accordingly. In terms of maternal utterances, only spontaneous speech was considered. Singing, non-verbal vocalizations and reading were left aside. Following guidelines discussed in Farran (2016), spontaneous utterances were subsequently labelled as adult-directed, infant-directed, or register-ambiguous. Utterances were thus judged intuitively (no acoustic analysis necessary), and classified as IDSp when including at least one of the following characteristics: pitch— or pitch range —notably exceeding that of typical adult-to-adult speech; long duration per syllable compared to adult-to-adult speech; smooth intonation with a soothing tone; sing-song pattern of rise and fall in intonation; parent production of infant vocalizations such as growls, squeals or raspberries; very long final syllables— longer than the lengthened final syllables used in adult speech to mark boundaries. All register-ambiguous utterances (23.2% of total maternal speech) were discarded.

Different routines make different pragmatic uses of language more or less probable (e.g. changing clothes makes conflict more probable than joint reading). As a consequence, different routines also tend to convey different prosodic profiles (e.g. fighting tends to entail louder volume than reading does). Because parents did not always record the exact same routine (e.g. playing, changing clothes, etc.), I-D utterances were labelled as either Inquiring, Gratifying, Pedagogic or Playful. Such a classification allowed for an equivalent prosodic comparison throughout the weeks of the study (i.e. only comparing gratifying IDSp sentences to each other). All utterances extracted from the same audio file and corresponding to the same category (e.g. all inquiring infant-directed utterances from the fourth week's audio file) were concatenated into a single audio file by means of a custom Praat script[23]. Because recording contained spontaneous speech, the presence of every single pragmatic function could not be guaranteed to take place every week (see Appendix H).

2.3.3.2. Analysis

Prosodic data was analysed by means of the Praat meta-script Prosogram (Mertens, 2004). Analysed f0 measures (fundamental frequency, see chapter 5 section 1) correspond to Mean, Median, Standard deviation, Maximum and Range, all of them measured in Hertz. Further measures included speech rate, the number of f0 contour glissandi, rises and falls, as well as the normalised Pairwise Variability Index (nPVI) of syllable nuclei durations.

[23] See acknowledgments.

Prosogram calculates speech rate as the number of nuclei divided by the sum of the total nuclei durations and the total internuclear durations, in a given file. Glissandi correspond to the proportion of nuclei with absolute f0 changes equal or larger than four semitones. Rises correspond to the proportion of nuclei with pitch changes equal or larger than four ascending semitones, and Falls to the equivalent regarding four descending semitones.

The nPVI is a measure of the average variation (or contrast) of a set of distances (in this case, time durations) that are obtained from successive adjacent ordered pairs of events (in this case, syllable nuclei) (Grabe & Low, 2002; Nolan & Asu, 2009). In very simple terms, this measure reflects the degree of difference between the durations of vowels in adjacent syllables. Given that IDSp has been described as rhythmically-enhanced, it can be expected that its nuclei durations will be more even and therefore its nPVI lower than that of ADSp (Lee et al., 2014).

Beyond analyses performed through Prosogram and Following Farran (2016), IDSp utterances per minute were calculated by taking the total duration of a given all IDSp concatenated file (e.g. all IDSp from the first week of the study) and dividing it by number of utterances. One utterance corresponded to one breath group, as implemented by Farran and recommended by Lynch and collaborators (Lynch et al., 1995). A ratio between the total number of IDSp utterances and the total number of ADSp utterances in a given week's recording was calculated (IDSp/ADSp). Finally, taking into consideration the role of IDSp as an ostensive cue discussed in chapters 3 and 6, the percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function) was also measured. A complete and detailed list of each of these measurements can be found in Appendix H.

2.3.4. Assessment of children's attention to recorded music

2.3.4.1. One-way remote video recording technique (ORViRT)

As reminded in this chapter's introduction, for methodological reasons the present study ought to generate instances in which children could— but would not be biased towards —autonomously engage with recorded music. Literature converges in suggesting that, along with singing, listening to music at home is one the most frequent musical activities as reported by parents— in Western societies (Custodero et al., 2003; Fancourt & Perkins, 2017; Ilari, 2005). Therefore, the aim was to gain access to such a natural and intimate setting, observing yet introducing as few unusual elements as possible.

To this end, a series of pilots were undertaken, aiming to generate a setting that would grant such an access. In them, participating parents were asked to engage in a Skype session were music would be played through internet hyperlinks provided by the researcher. The researcher would call the participant when the child was already in the same room. Participating parents joined the Skype meeting with their computers on mute mode so as not to draw the child's attention with the session's initial ringing, and communicated with the researcher through Skype's chat. The researcher had his camera and microphone off, so as not reveal his presence to the child. Through the chat, he fed the parents YouTube links to Disney song clips and nursery rhymes.

Out of these experiences, the need to introduce a number of amendments to the methodology was made clear. First, children being in the room with nothing occupying or distracting them by the time calls started constituted a problem. Although the call itself was silent, the very fact that parents would manipulate their laptop in front of their children made the latter repeatedly interested in doing the same. Because parents still tried to communicate with the researcher, quarrels concerning the use of the computer took place, impeding any music being played in the first place, or often ending the call because of buttons being pressed by the toddler. A first amendment thus consisted in parents being asked to prepare the setting alone, only afterwards bringing their children in, providing some source of entertaining for them, and only then starting the call.

Using the laptop both for communicating with the researcher and playing the music proved problematic beyond the mentioned quarrels. In such a scenario, from the moment the music started it was clear to the child that the parent's attention was focused in the computer. As a result, its attention was often drawn to it too, even if by mere curiosity or imitation. This made it in turn impossible for the researcher to judge whether it was the music being played through the laptop, or the parent's attention to the laptop itself, what was drawing the toddler's attention. Therefore, a second amendment was introduced. Laptops would be used to record the scene, but music would be played through a different device: the parent's smartphone. Such a device allowed for a more discrete manipulation, which seemed to distract children comparatively-less.

A similar issue was found regarding music clips containing moving images. Because screens and related visual stimuli are highly appealing to children (Hadlington et al., 2019), whenever internet hyperlinks led to music clips containing moving images, it was impossible to judge whether it was the music being played or the accompanying images, what was drawing the toddler's attention. For this reason, all clips used in the study contained no moving images.

As a result of the piloting, an amended methodology was used. Participating parents were asked to engage in a 15-minutes Skype (or Google Hangouts) session on weekly basis, the steps of which were as follow. 1) Prior to each meeting, participants would ready the setting: a room where the target child could play with some toys without being disturbed. The computer should be on mute mode so as not to draw the child's attention and placed so that the integrated camera would capture the child and its immediate surroundings. 2) The parent would bring the target child into the room and invite them to play on their own. 3) The parent would, only then, confirm to be ready via WhatsApp. At the researcher's confirmation, he or she would discretely start the call. Again, in order not to draw the child's attention, the researcher's microphone and camera were turned off. By this token, the child could not see or hear the researcher whereas the latter could do both and record the session by using QuickTime Pro record screen function. 4) During the session, the researcher would feed internet links to the parent via WhatsApp chat, leading to songs systematically selected and detailed in the next section. Parents' use of their smartphone was mainly limited to clicking on hyperlinks, only occasionally texting back the researcher, and never engaging in voice phone calls. Because screens and visual stimuli are highly appealing to children and in case the latter would direct their gaze at the device, links lead specifically to audio-only versions of songs. 5) Once all links corresponding to a particular session were played, the researcher ended the video call.

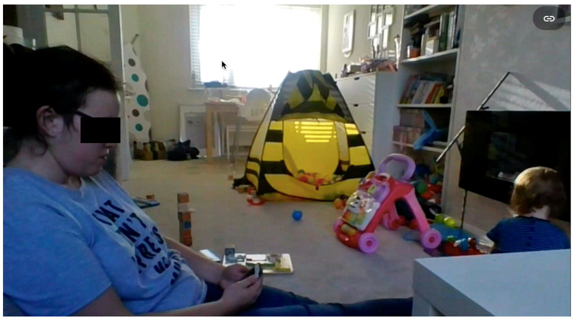


Figure 1a. Louisa playing on her own.

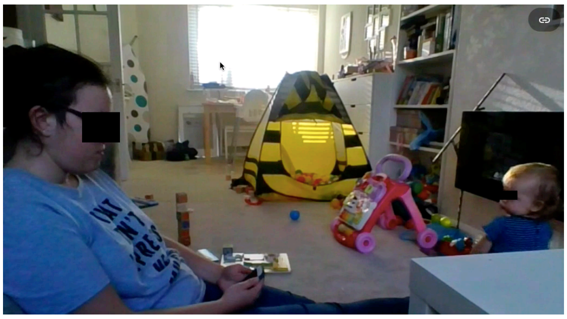


Figure 1b. Louisa's gaze (and attention) being directed at the source of the music. Starting point of the attentional bout.

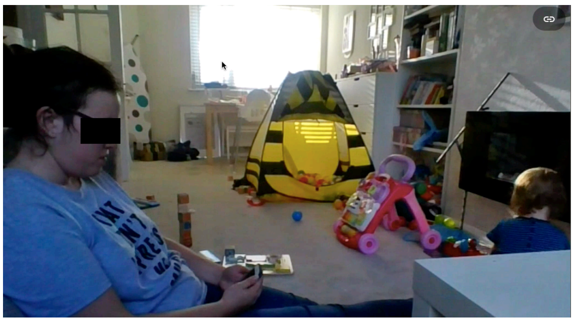


Figure 1c. Louisa losing interest in the music being played and returning to her toys. Ending point of the attentional bout.

Figures 1 a-c. Screenshots from an ORViRT session. Figure 1a presents a participating child playing on her own as she normally would, completely unaware of the recording situation. Figure 1b shows the moment in which the participating mother plays on her phone music fed by the researcher. The child's attention is drawn to the source of the music, as evidenced by her gaze. After a few seconds (two in this particular example), the child's attention shifts back to her toys, thus ending the attentional bout (Figure 1c). The few-seconds attentional bout thus represents an instance of initial fixation.

As a result, each session started with the child playing as it normally would, in an ecologically-valid environment (Figure 1a). Only then, music would be played by the parents and it was up to the child to pay attention to it according to their spontaneous interest, or not (Figure 1b). A crucial element of the devised setting is that parents were instructed to remain neutral to the sound of music, refraining from prompting music-listening or any other particular reaction, so at to allow children's spontaneous reactions to take place. At the same time, they were free to respond as they normally would, should the child spontaneously engage them in any way as a reaction to the music being played. They were thus allowed to return gazes and smiles, dance or do actions *if* requested to.

2.3.4.2. Musical stimuli

Participating children were exposed to music that varied systematically in terms of genre, familiarity, and history of interaction (the complete list can be found in Appendix I). Following conclusions from the second study's discussion, participating children were exposed to only two genres: either nursery rhymes and children's songs, or pop. All songs were played through the participating mothers' smartphones and provided via links to Soundcloud music tracks (www.soundcloud.com). Soundcloud was chosen because it provided versions of the songs that comprised no moving images. Whenever a song was not available in Soundcloud, a YouTube link (www.youtube.com) was provided instead, also making sure to consist of sound-only tracks. The complete list of internet hyperlinks used in the study can also be found in Appendix I.

After checking during preliminary informal conversations that all children were familiar with it, the traditional British song 'Twinkle-twinkle little star' was chosen as a familiar nursery rhyme (FamNurs) and played in every session. 'Every little thing' by The

Police was played during the first three weeks of the study and considered as familiar pop (FamPop) from the fourth session onwards.

A pool of foreign nursery rhymes and children's songs were also presented to participating children (UnfamNurs). In order to guarantee the unknown character of the pieces, they were sung in foreign languages (French and Spanish) and never presented twice during the study. Additionally, mothers were asked to notify the researcher in case they happened to have previously exposed their children to any of them (it was never the case).

As hinted in this chapter's introduction, a further goal of the present study was to isolate and assess the importance of social interaction. To this effect, participating parents were asked to interact with their children to the sound of one particular nursery rhyme or children's song of their choice. Instructions simply invited them to enjoy themselves together to the sound of music, once a week. Participants 4-6 chose 'The wheels on the bus' and participants 1 and 2 preferred 'If you're happy and you know it'. In order to maximize their natural character and to avoid tension, such interactions did not take place during the recorded sessions, but at some other moment during the week thus without being observed by the researcher. Informal WhatsApp chats with participants revealed that they mainly consisted of doing the actions associated with these songs and 'goofin' around'. Participants did this during the first seven weeks of the study. From the eighth week onwards, each participant's chosen nursery rhyme or children's song was played once during each ORViRT session, guaranteeing that children would have associated it to pleasurable social interaction with a significant other. At the same time, from the eighth week onwards, the weekly off-camera interactions were discontinued. Such pieces were thus labelled as Interactive Nursery Rhymes (InterNurs). In a similar fashion, participating parents were asked to interact with their children off-camera using one particular pop song of their choice that was deer to them. This time, instructions invited them to show to their children, in whatever way they deemed appropriate, that the chosen songs had meaning for them. Participants— except for participant number 5, who felt unable to choose a song —also did this during the first seven weeks of the study, their song being played once during each session from the eighth week onwards and labelled as Interactive Pop (Inter Pop). Taken together, the kind of musical pieces discussed in this paragraph will be referred to as ‘interactional songs’.

2.3.4.3. Video analysis

Recorded sessions were saved as .MOV files and annotated by means of the software ELAN (2019). Attentional bouts were measured primarily through the child’s visual fixation, their start and ending points being annotated in a tier and the resulting length of the attentional bout measured in seconds. Occasionally, children would also freeze their playing, clearly showing that their attention was focused on a song.

All ambiguous signs of attention were discarded from analysis. For instance, during some sessions children would not only direct their gaze at the mother’s smartphone (the source of music) but also insistently attempted to manipulate it. As a consequence, in such cases it was not possible to distinguish between the child’s interest in the music being played and an apparently competing interest in the device itself.

Although it was not contemplated at the time of participant recruitment as a proper assessment tool, at the end of the study participants were requested to participate in a short, informal post-hoc interview during which the researcher could collate qualitative impressions gathered during the study, as well as debriefing them regarding the aims and preliminary results of the study. Four interviews were performed, recorded and transcribed, and can be found in Appendix J. Participants 4 and 6 did not show an interest in participating and were not interviewed.

2.4. Hypotheses

2.4.1. Distinct, 'quantal' score increases between fortnightly LUI scores will be followed by distinct changes in the mothers' IDSp features (shifting towards A-D prosodic standards).

2.4.2. There will be significant, negative correlations between LUI raw scores and the mother's IDSp features (contained in the weekly audio files). As LUI scores are expected to increase over time, the features of IDSp prosody are expected to shift towards A-D prosodic standards[24].

2.4.3. There will be significant, negative correlations between the mothers' IDSp features and the length of attentional bouts that children pay to recorded music (as evidenced through the ORViRT sessions). As IDSp shifts towards A-D prosodic standards, bouts will become increasingly longer.

2.4.4. Whenever hypothesis 2.4.1 proves correct, distinct changes in the mother's IDSp prosody will be followed by distinct increases in terms of the length of the attentional bouts that children pay to recorded music.

2.4.5. There will be inter-group differences in terms of the mother's IDSp prosody

2.4.6. There will be inter-group differences in terms of the children's prosody

2.4.7. There will be inter-group differences in terms of the length of the attentional bouts that children pay to recorded music

[24] Because the assessed prosodic features have been precisely defined in section 2.3.3, it is possible to generate equally-precise sub-hypotheses. Concerning f0, changes were expected in one or all measures. More concretely, it was expected that these measures would lower (e.g. lower mean in Hz, smaller ranges, etc.) as a consequence of changes in the LUI scores. F0 contour glissandos, rises and falls were expected to diminish in number (e.g. less glissandos will be detected). In terms of nPVI and Speech Rate, and IDSp utterances per minute, measures were expected to increase over time. Ratios between the total number of IDSp utterances and the total number of ADSp utterances were expected to decrease.

3. Results

3.1 Longitudinal case studies

In this section, the case of each participating dyad will be assessed both quantitively and qualitatively, enabling the assessment of hypotheses 2.4.1-4. Concerning quantitative data, in each case, data will be presented as follows. First, LUI scores will be presented and the presence of quantal changes will be examined, providing— or not —grounds for the assessment of hypothesis 2.4.1. Next, data corresponding to the initial questionnaire and prosodic data will be summarized. In order to asses hypothesis 2.4.2, all correlations between LUI and prosodic data will be presented. Whenever LUI score quantal changes had been found, an assessment of hypothesis 2.4.1 will take place. Because the amount of data is already generous and in order to be efficient, data will not necessarily be plotted identically across participants. A rather adhocratic approach will instead be used, privileging the plotting of data that proves relevant. Next, attentional data will be reported. Once again, in order to assess hypothesis 2.4.3, all correlations between prosodic and attentional data will be presented and, should prosodic quantal changes be found, assessment of hypothesis 2.4.4 will take place.

Following the presentation of quantitative data, a qualitative appraisal of the case will follow. Such an appraisal will provide the researcher's impressions concerning the child's particularities, including general behaviour, exchanges with the caregiver, and success of the observational setting. Equally importantly, this appraisal will attempt a first integration of the different kinds of quantitative data already presented, fed by further information about the child and its family environment as reported in the initial questionnaire and post-hoc interviews.

3.1.1. Hugh[25]

3.1.1.1. Initial questionnaire, LUI, vocal and attentional data

Boy, 15.9 months-old when started the study. His weekly LUI raw global scores show a slow but constant increase in his pragmatic skills, as corroborated by a Pearson's Linear Least Squares Regression (R^2 = .82) (Figure 6). A soon as Hugh turned 18 months old and his LUI raw scores could be compared to the rest of his age peers (the last four questionnaires), he positioned himself consistently in the 100th percentile. Two distinct score rises can be observed, one at week 7, and one at week 13 (Figure 2).

Question	Score
1.1 My child understands everything I say to they	1
1.2 My child understands sentences	2
1.3 Sometimes, my child understands me, but pretends otherwise	-2
1.4 Sometimes, my child does not cooperate because they do not understand what I'm asking for	-1
1.5 My child has learned to walk	3
2 How autonomous would you say your child is?	5

Table 3. Initial questions concerning the mother's appraisal of Hugh's development and autonomy. Questions 1.1-5 correspond to Likert scales where -3 = I do not agree at all, and 3 = I completely agree. Question 2 corresponds to a Likert scale where 0 = Very dependent, and 10 = Very autonomous.

In order to assess hypothesis 2.4.2, prosodic measures extracted weekly from recordings (see section 2.3.3 in this chapter) were correlated with their corresponding LUI raw scores. In order to be preserve weekly comparisons and because LUI scores were provided only fortnightly, an average was calculated between weeks (i.e. if the first week's LUI raw score corresponded to 40 and the third week to 42, the second week gap was filled with a corresponding 41). Table 4 displays the correlation coefficients corresponding to the 73 prosodic items that result of the combination of the different speakers, speech registers and pragmatic functions described in section 2.3.3.2. Amid these 73 elements, 11 significant correlations were found. IDSp (not divided into any particular pragmatic function) seems to

[25] The names of the children have been changed in order to protect their identity.

concentrate the majority of significant correlations. However, in terms of f0-related measures, significant correlations are positive (i.e. increased along with LUI scores), thus contrary to hypothesis 2.4.2. The same applies to Playful IDSp f0 median. Something similar can be observed in terms of IDSp utterances per minute, which were expected to increase. On the other hand, nPVi and IDSp/ADSp utterances ratios behaved as expected. The former increased along with LUI scores, as the latter decreased.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	-0.2	-0.4	-0.4	-0.4	-0.4	-0.4	0.1	-0.2	0.2	0.5*		
Ch	/	-0.1	-0.2	-0.3	-0.3	-0.3	0.3	0.0	0.3	/		
ID	0.3	0.4	0.4	0.5	0.5*	0.7*	0.5**	0.4*	0.4*	0.3*	-0.5*	-0.7**
Inq	-0.2	-0.1	-0.2	-0.2	-0.1	0.1	0.3	-0.1	0.3	-0.2		
Grat	0.0	0.1	0.1	0.2	0.1	0.1	0.2	0.3	0.1	0.0		
Ped	0.0	-0.2	-0.2	0.1	0.1	0.1	0.3	0.4	0.3	-0.1		
Play	0.0	0.7**	0.6*	-0.1	0.0	0.2	0.4	0.1	0.4	0.2		

* = p < .05, ** = < p < .01

Table 4. Pearson correlations coefficients between LUI raw scores and prosodic measures per prosodic item, encompassing speaker, speech register and pragmatic function (see section 2.3.3 in this chapter). Columns represent prosodic measures: Speech rate (SpR), f0 Median in Herts (Med), f0Mean (Hz) (Mn), f0 Standard deviation (SD), f0 Maximum (Hz) (Mx), f0 Range (Hz) (Rg), f0 Glissandi (Gls), f0 Raises (Rss), f0 Falls (Fls), normalised Pairwise Variability Index of syllable nuclei durations (nPVI), IDSp/ADSp utterances ratio (ID/AD) IDSp utterances per minute (Ut/m), and (%ID). Speech rate and nPVI are missing given that children's vocalizations often consisted of non-verbal expressions, out of which these two measures cannot be calculated. Rows represent speakers, speech registers and pragmatic functions. In terms of speakers, Hugh's vocalizations are labelled as 'Ch', and all the rest correspond to his mother's utterances. In terms of register, maternal utterances are labelled as adult-directed ('AD') or infant-directed ('ID'), the latter further divided into four pragmatic functions: Inquiring IDSp (Inq), Gratifying IDSp (Grat), Pedagogic IDSp (Ped) and Playful IDSp (Play).

The mother's IDSp Speech Rate does evidence a sudden increase in week 9 (thus following the LUI increment on week 7). As discussed, ADSp speech has higher Speech Rate than IDSp, so such a change would point in the direction of Hypothesis 1.4.1. However, a higher Speech Rate is not maintained throughout the following weeks, nor it is echoed by Table 4, and thus cannot be interpreted as support for the hypothesis. A similar pattern can be observed in the mother's f0 Range (Figure 2), which does evidence a sudden rise on week 8 (again, thus following the LUI raise on week 7) this time supported by a significant correlation in Table 4. Still, even though such a distinct prosodic change follows a distinct LUI score increase, it constitutes a shift away from A-D prosodic standards and does not finally support Hypothesis 2.4.1. The lack of support for this hypothesis renders, in turn, the assessment of hypothesis 2.4.4 (distinct changes in attentional bout duration following distinct changes in parental prosody) impossible.

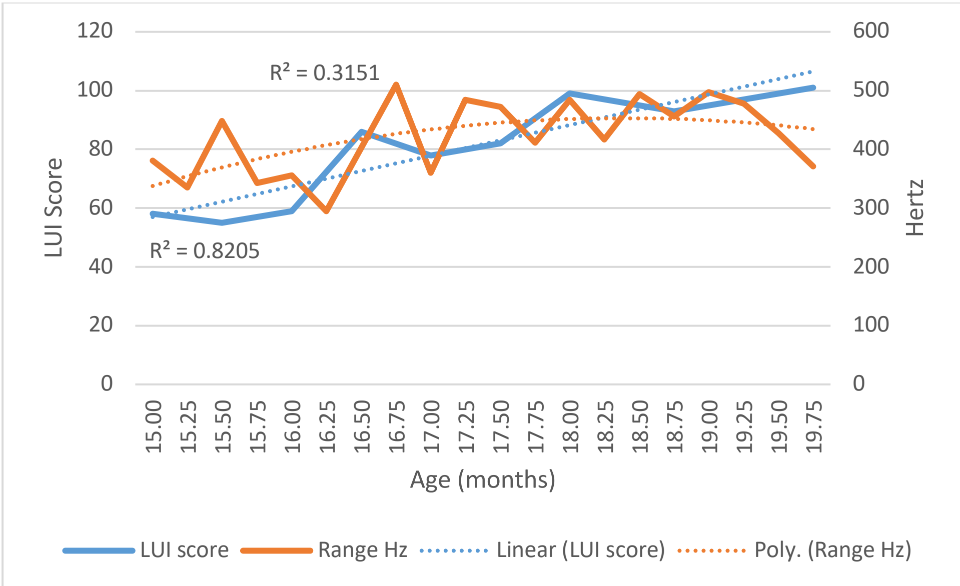


Figure 2. Upper left-hand side. Hugh's LUI scores (blue line) and Hugh's mother's f0 Range in hertz (orange line), per month of age. The dotted blue line represents a fitted linear regression model. The orange dotted line represents a fitted polynomial regression model.

In order to asses hypotheses 2.4.3, the next step was to introduce attentional data. As previously discussed (section 2.3.4.2 in this chapter), in an attempt to isolate and assess the importance of social interaction, participating parents were asked to interact off-camera with their children to the sound of a nursery rhyme and pop song of their choice (InterNurs and InterPop, respectively). From the eighth week onwards, such songs were played once during each ORViRT session, while being discontinued off-camera. Figure 3 displays the average attentional bouts paid by Hugh to musical stimuli, both including and excluding interactional songs. As reflected by the practically identical fitted linear models, including or excluding interactional songs does not change the overall pattern: attentional bouts decrease as time progresses, with a peak around the introduction of interactional songs.

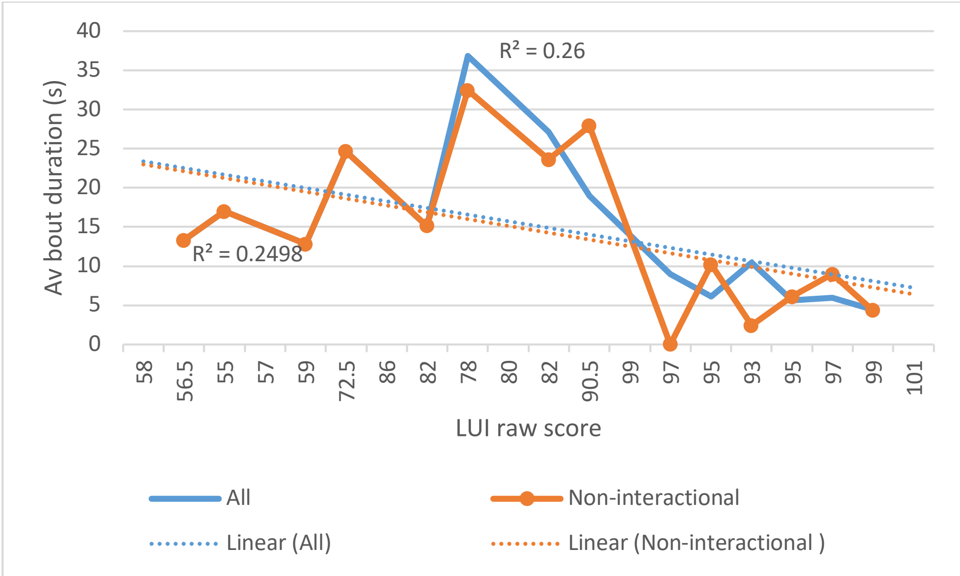


Figure 3. Lower left-hand side. Hugh's average attentional bout duration to all musical stimuli, in seconds (blue line), and average attentional bouts to all musical stimuli except those involving interaction (InterNurs and InterPop), in seconds (orange line), per LUI raw score. Dotted lines represent corresponding fitted linear regression models.

This is also reflected in Table 5, where the overall attentional bout length average, along with familiar pop, unfamiliar nurseries, and interactive nurseries evidence significant negative correlations with LUI raw scores. Interactional songs seem to consistently increase averages; when considered (blue line), interactional songs do seem do slightly increase averages, as seen in datapoints corresponding to LUI scores 78, 97 and 93. The fact that attentional bouts tend to decrease, directly contradicts hypotheses 1.4.3 and 2.4.4, both of which predicted increments in terms of attentional bout length. No significant patterns from week to week were obtained when separated into the different genres (nursery rhymes/pop), degrees of familiarity (familiar/non familiar) or interaction (interactive/non interactive).

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
Correlation	-0.3	-0.6*	-0.8**	-0.5	-0.8**	-	-0.6*	-0.6

* = p < .05, ** = < p < .01

Table 5. Pearson correlation coefficients between LUI raw scores and average attentional bouts, per genre, familiarity, and whether were associated with interaction or not (see section 2.3.4.2 in this chapter). Columns thus correspond to familiar nurseries (FamNurs), familiar pop (FamPop), unfamiliar nurseries (UnfamNurs), unfamiliar pop (UnfamPop), nurseries associated with social interaction (InterNurs), and pop associated with social interaction (InterPop). The average of the previous stimuli is also presented (Av), as well as the average of the first four stimuli, not associated with interaction (Av-Inter). No value is shown in the cell corresponding to InterPop, as Hugh paid no attention whatsoever to such musical stimulus.

Still, correlations coefficients between prosodic measures and average attentional bouts excluding interactional songs (the main interest of this dissertation) are presented in Table 6. The 10 correlations found suggest only partial, unsystematic relationships between parental prosody and average attentional bout duration. Finally, and outside any particular hypothesis, correlations were also drawn between LUI raw scores and attentional bout lengths (Table 6).

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	0.5*	0.3	0.4	0.3	0.3	0.3	-0.1	0.1	-0.1	-0.5*		
Ch	/	0.1	0.2	0.2	0.1	0.0	-0.4	0.2	-0.5*	/		
ID	0.3	-0.1	0.0	0.0	0.1	-0.1	-0.6**	-0.5*	-0.3	-0.3	-0.2	0.2
Inq	0.2	0.1	0.2	0.4	0.6**	0.5*	-0.5*	-0.4	-0.3	-0.4		
Grat	0.2	0.1	0.1	0.1	0.2	0.2	0.0	-0.2	0.1	-0.5		
Ped	0.2	0.5	0.6**	0.2	0.4	0.3	-0.3	-0.3	-0.3	0.2		
Play	0.4	-0.5	-0.3	0.4	0.2	-0.5*	0.0	0.2	-0.1	-0.1		

* = p < .05, ** = < p < .01

Table 6. Pearson correlation coefficients between prosodic measures and average attentional bouts excluding interactional songs per prosodic item, encompassing speaker, speech register and pragmatic function (see section 2.3.3 in this chapter). Columns represent prosodic measures: Speech rate (SpR), f0 Median in Herts (Med), f0Mean (Hz) (Mn), f0 Standard deviation (SD), f0 Maximum (Hz) (Mx), f0 Range (Hz) (Rg), f0 Glissandi (Gls), f0 Raises (Rss), f0 Falls (Fls), normalised Pairwise Variability Index of syllable nuclei durations (nPVI), IDSp/ADSp utterances ratio (ID/AD), IDSp utterances per minute (Ut/m), and (%ID). Rows represent speakers, speech registers and pragmatic functions. In terms of speakers, Hugh's vocalizations are labelled as 'Ch', and all the rest correspond to his mother's utterances. In terms of register, maternal utterances are labelled as adult-directed ('AD') or infant-directed ('ID'), the latter further divided into four pragmatic functions: Inquiring IDSp (Inq), Gratifying IDSp (Grat), Pedagogic IDSp (Ped) and Playful IDSp (Play).

3.1.1.2. Qualitative appraisal

Hugh is a little boy already able to walk by the study's start date. His LUI scores are higher than the other two participants in the study, and his percentiles theoretically place him above the vast majority of his contemporaries. These findings can be at least partially explained by the fact that Hugh is the only participating child that has older siblings— two of them, see Table 3 in the previous section. Unlike first-time mothers, that have comparatively-less experience and skill, Hugh's mum is more experienced when it comes to perceiving her children's incipient speech. Thus, surrounded by more linguistically-proficient siblings and a more experienced mother, Hugh's pragmatic development is not surprising. In this context, the fact that both his mother's rate of IDSp utterances per minute and IDSp/ADSp ratio decreased over time is interesting. It suggests Hugh's mother has entered a phase where she was using comparatively-less IDSp, and the quality of IDSp itself was changing.

Al sessions took place in his playroom and started with him playing autonomously next to his mum. His engagement with toys gave the impression of a somewhat short attention span. Yet, out from the first session, he proved being capable of paying sustained attention to music, on his own, as he did as a reaction to some nursery rhymes. Also from the beginning, he gave signs of understating words (lyrics), by performing corresponding actions. Throughout the sessions, and as soon as music started coming out of him mum's smartphone, Hugh actively and rather insistently tried to reach for and manipulate it. During the first three sessions, he showed repeated signs of frustration when not being granted access to the device, eventually throwing tantrums. When asked about her son's interest in the smartphone (see appendix J), the mother explained that at the time of the study her and her partner tried not to let him near screens. Her interpretation was that Hugh was fascinated with them precisely because they were new and interesting.

His mum's reaction to these tantrums consisted in semi-humorous comfort phrases like 'oh Bill...' uttered in IDReg, followed by cuddles. She also seemed rather used to drawing his attention quite actively when in need to sooth him. When compared to the other mothers in this age group, she also seemed comparatively-hastier when it came to deal with her son's distress. This seems contrary to her own rating of Hugh's autonomy— also higher than the rest in this age group —which, combined with the fact that this is not a first-time mother, rather suggested that she could have been more relaxed. Her hastiness could thus rather be attributed to anxieties related to being knowingly observed. This was confirmed by a post-hoc interview (see Appendix J), where she acknowledged feeling more natural only after the first few sessions. It is thus likely that the combination of the mother's initial distress plus the frustration of being denied access to the smartphone led Hugh to the tantrums.

From the fourth session onwards, and after exchanges with the researcher, his mum changed the strategy regarding the device. Instead of keeping it away from Hugh whatever the cost, she opted for distracting him and talking to him, or simply granting him access to it. From week eight onwards (around 16.75 months of age), a new pattern seemed to emerge as a response to music: he would dance or make song-appropriate actions on his own (seeking for less visual contact) for rather short periods of time— literally one or two seconds. The ephemeral quality of these movements made them look somewhat stereotypical, as if he wanted to demonstrate he knew that to do when faced with music, rather than taking some time to actually enjoy himself through it. Such behaviours would sometimes (particularly around week 9, 17 months of age) occur in parallel to longer attentional bouts than before. From around weeks 12-14, new changes in his behaviour seemed to take place: he looked more focused and calmer. At the same time, he paid less attention to music in general, as evidenced in Figure 3. Such changes can be interpreted in at least two ways. They could have reflected that his concentration capacity augmented, thus being less 'distracted' by music that was played while he was busy playing with his toys in the first place. Alternatively, Hugh had perhaps has learned by then that a large extent of what came from trying to reach mum's smartphone was fighting with her over it.

3.1.2. Louisa

3.1.2.1. LUI, vocal and attentional data

Girl, 15.2 months-old when started the study. Her LUI scores show, as expected, a slow but constant increase in her pragmatic skills, as corroborated by a Pearson's Linear Least Squares Regression (R^2 =.94). As Louisa turned 18-months-old, her LUI raw scores could be compared to the rest of her age peers. In those last three questionnaires, she located herself twice in the 1st percentile and once in the 4th. No noticeable score increases can be observed (Figure 4), making the assessment of hypothesis 2.4.1 and 2.4.2 impossible.

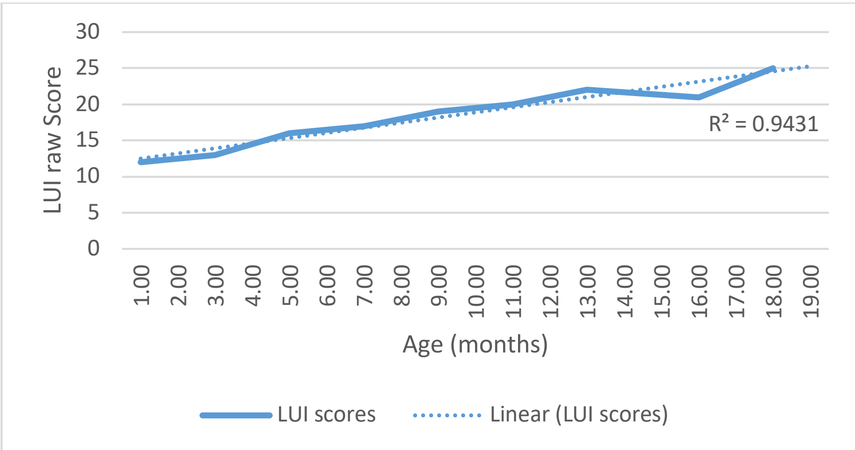


Figure 4. Luisa's LUI scores per week of age. The dotted line represents the R^2 fitted line

Question	Score
1.1 My child understands everything I say to they	-2
1.2 My child understands sentences	-2
1.3 Sometimes, my child understands me, but pretends otherwise	-1
1.4 Sometimes, my child does not cooperate because they do not understand what I'm asking for	2
1.5 My child has learned to walk	-2
2 How autonomous would you say your child is?	2

Table 7. Initial questions concerning the mother's appraisal of Louisa's development and autonomy. Questions 1.1-5 correspond to Likert scales where -3 = I do not agree at all, and 3 = I completely agree. Question 2 corresponds to a Likert scale where 0 = Very dependent, and 10 = Very autonomous.

Seven, rather isolated prosodic measures in the voice of Louisa's mother correlated significantly with LUI raw scores. Because the lack of noticeable score increases disabled the assessment of hypothesis 1.4.1, no prosodic measures are plotted. Again, support for hypothesis 2.4.2 is only partial. In support, there are positive correlations between LUI raw scores and AD speech rate, IDSp and inquiring IDSp nPVIs, as well as negative correlations with pedagogic IDSp f0 median and mean. On the other hand, a negative correlation with pedagogic IDSp speech rate along with positive correlations with the number of f0 rises in gratifying and playful IDSp stand as evidence against it.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	0.5*	0.4	0.3	0.2	0.2	0.2	0.3	0.3	0.2	0.1		
Ch	/	0.0	0.0	-0.2	-0.2	-0.2	0.2	0.0	0.3	/		
ID	0.1	-0.3	-0.1	0.2	0.2	0.2	0.0	0.2	-0.1	0.5*	0.4	0.2
Inq	-0.2	-0.4	-0.3	-0.1	-0.1	0.0	-0.2	-0.3	0.2	0.5*		
Grat	-0.1	0.0	0.0	0.0	0.0	0.0	0.5	0.6*	0.2	0.3		
Ped	-0.4	-0.6*	-0.5*	-0.3	-0.3	-0.2	0.3	0.2	0.3	-0.2		
Play	0.1	-0.1	0.0	-0.1	0.0	0.0	-0.1	0.5*	-0.1	-0.4		

* = p < .05, ** = < p < .01

Table 8. Pearson correlation coefficients between LUI raw scores and prosodic measures, per speaker, speech register and pragmatic function (see section 2.3.3 in this chapter). Columns represent prosodic measures: Speech rate (SpR), f0 Median in Herts (Med), f0Mean (Hz) (Mn), f0 Standard deviation (SD), f0 Maximum (Hz) (Mx), f0 Range (Hz) (Rg), f0 Glissandi (Gls), f0 Raises (Rss), f0 Falls (Fls), normalised Pairwise Variability Index of syllable nuclei durations (nPVI), IDSp/ADSp utterances ratio (ID/AD), IDSp utterances per minute (Ut/m), and (%ID). Speech rate and nPVI are missing given that children's vocalizations often consisted of non-verbal expressions, out of which these two measures cannot be calculated. Rows represent speakers, speech registers and pragmatic functions. In terms of speakers, Louisa's vocalizations are labelled as 'Ch', and all the rest correspond to her mother's utterances. In terms of register, maternal utterances are labelled as adult-directed ('AD') or infant-directed ('ID'), the latter further divided into four pragmatic functions: Inquiring IDSp (Inq), Gratifying IDSp (Grat), Pedagogic IDSp (Ped) and Playful IDSp (Play).

Figure 5 displays the average attentional bouts paid by Louisa to musical stimuli, both including and excluding interactional songs. Again, both fitted linear models present the same pattern, which this time consists in an overall increment. Also as in the previous case, a strong peak around the introduction of interactional songs can be observed. The size of such a sudden increment is noticeably less prominent when excluding interactional songs (orange line in the graph). This is also the case in further datapoints corresponding to LUI scores 20.3 and 21. As in the previous case, correlations were also drawn between LUI raw scores and attentional bout lengths. This time, only the average excluding interactional songs returned a significant (positive) correlation (table 9). No significant patterns from week to week were obtained when separated into the different genres (nursery rhymes/pop), degrees of familiarity (familiar/non familiar) or interaction (interactive/non interactive).

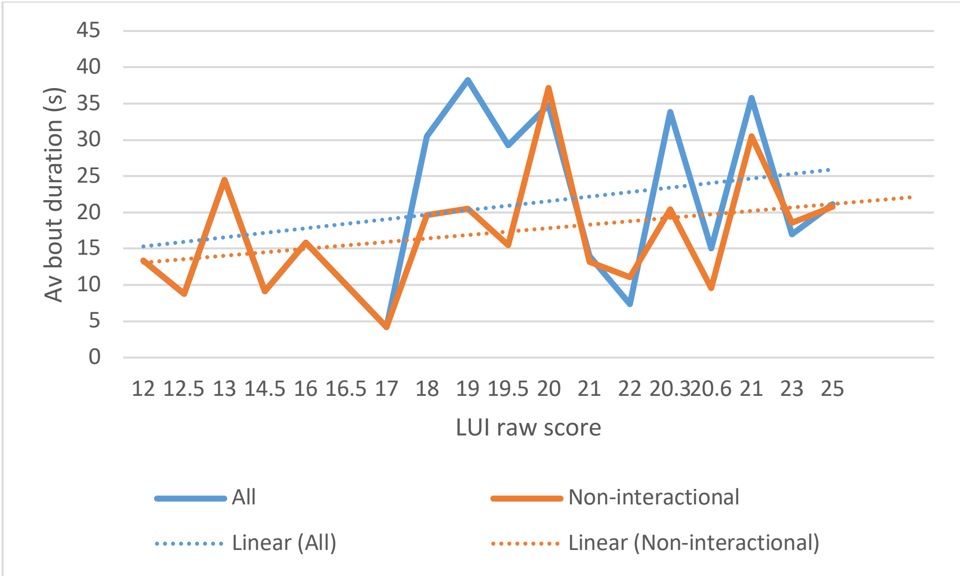


Figure 5. Lower left-hand side. Louisa's average attentional bout duration to all musical stimuli, in seconds (blue line), and average attentional bouts to all musical stimuli except those involving interaction (InterNurs and InterPop), in seconds (orange line), per LUI raw score. Dotted lines represent corresponding fitted linear regression models.

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
Correlation	-0.1	0.5	0.3	-0.3	-0.4	-0.3	0.4	0.4*

* = p < .05, ** = < p < .01

Table 9. Pearson correlation coefficients between LUI raw scores and average attentional bouts, per genre, familiarity, and whether were associated with interaction or not (see section 2.3.4.2 in this chapter). Columns thus correspond to familiar nurseries (FamNurs), familiar pop (FamPop), unfamiliar nurseries (UnfamNurs), unfamiliar pop (UnfamPop), nurseries associated with social interaction (InterNurs), and pop associated with social interaction (InterPop). The average of the previous stimuli is also presented (Av), as well as the average of the first four stimuli, not associated with interaction (Av-Inter).

Correlations coefficients between prosodic measures and average attentional bouts excluding interactional songs are presented in Table 10. A total of 9 correlations reached statistical significance. Most f0-related measures corresponding to pedagogic IDSp present negative correlations with the evolution of average attentional bout durations, which stands as support for hypothesis 2.4.3.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	-0.2	0.4	0.4	0.3	0.4	0.4	0.2	0.1	0.2	-0.2		
Ch	/	0.5*	0.3	0.2	0.3	0.2	0.1	0.2	0.1	/		
ID	-0.1	0.2	0.1	0.1	0.1	0.1	-0.2	-0.4	0.0	-0.1	-0.3	-0.1
Inq	0.4	-0.1	-0.2	-0.3	-0.3	-0.4	0.8**	0.8**	0.1	-0.2		
Grat	-0.2	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.2	0.0	0.7*		
Ped	0.2	-0.6**	-0.6*	-0.5*	-0.6**	-0.6**	-0.3	-0.3	-0.3	0.0		
Play	0.1	-0.1	0.0	0.2	0.2	0.2	0.0	-0.1	0.0	-0.1		

* = p < .05, ** = < p < .01

Table 10. Pearson correlation coefficients between prosodic measures and average attentional bouts excluding interactional songs, per speaker, speech register and pragmatic function (see section 2.3.3 in this chapter). Columns represent prosodic measures: Speech rate (SpR), f0 Median in Herts (Med), f0Mean (Hz) (Mn), f0 Standard deviation (SD), f0 Maximum (Hz) (Mx), f0 Range (Hz) (Rg), f0 Glissandi (Gls), f0 Raises (Rss), f0 Falls (Fls), normalised Pairwise Variability Index of syllable nuclei durations (nPVI), IDSp/ADSp utterances ratio (ID/AD), IDSp utterances per minute (Ut/m), and (%ID). Rows represent speakers, speech registers and pragmatic functions. In terms of speakers, Louisa's vocalizations are labelled as 'Ch', and all the rest correspond to her mother's utterances. In terms of register, maternal utterances are labelled as adult-directed ('AD') or infant-directed ('ID'), the latter further divided into four pragmatic functions: Inquiring IDSp (Inq), Gratifying IDSp (Grat), Pedagogic IDSp (Ped) and Playful IDSp (Play).

3.1.2.2. Qualitative appraisal

By the study's start date, Louisa was not walking yet, but crawled sitting on her bottom. This confirms figures in Table 7, which suggest that her mum's appraisal in terms of Louisa's autonomy and development was comparatively-lower than the other two children in this age group. All sessions took place in her playroom and started with her playing autonomously at a certain distance from her mum. Louisa looked, in general calmer than, for instance, Hugh. Nevertheless, she also threw a tantrum in the first session, to which her mother reacted calmly, cuddling her for a moment, and providing further toys as a distraction. Until around week 9, whenever her daughter reached her as a reaction to music Louisa's mum often looked somewhat rigid in her posture, smiled back only slightly, and offered little vocalization compared to all other participating mothers. This gave the impression of her feeling inhibited, which she confirmed in a post-hoc interview (appendix J). Because the study's instructions did not prevent her from reacting to her daughter's spontaneous reactions to music, she was reminded of her freedom to do so.

Until around week 7, Louisa did not seem to be that often interested in music, perhaps partially as a result of her mother's comparatively-low responsivity, which seemed to lower the chances of the toddler paying sustained attention to music. On the other hand, from week 9 (shortly after the introduction of interactional songs) onwards such a pattern seems to change. Louisa starts looking for eye contact sometimes up to the entire length of a song. Her mum starts smiling back rather shyly, as if by means of emotional contagion. They also seem to start sharing a look of complicity. Louisa' enthusiasm for music seems to be more evident on camera: she dances in rhythm more enthusiastically and engages in more overt signs of affection towards her mum, like little kisses. Such enthusiasm was particularly evident to the sound of 'If you're happy and you know it', the nursery rhyme that had been chosen for previous interaction (Inter Nurs). This impression is supported by the notorious increment in terms of attentional bout length between that interactional songs introduced in the case of Louisa, as presented in Figure 5 and discussed in the paragraph above it. The mentioned enthusiasm is also remarkable considering that this dyad did not have a particular relationship with this song prior to the study (see post-hoc interview), suggesting that Louisa was particularly receptive to the song and the associated interaction that had taken place off-camera. After week 13, some of these reactions seem to wane in their intensity. Attention also seems to be progressively selective towards the song's lyrics.

3.1.3. Daisy

3.1.3.1. LUI, vocal and attentional data

Girl, 16.6 months-old when started the study. Her LUI scores show, as expected, a slow but constant increase in her pragmatic skills, as corroborated by a Pearson's Linear Least Squares Regression (R^2 = .90). As Daisy turned 18-months-old, her LUI raw scores could be compared to the rest of her age peers. In those six last questionnaires, the percentile she located herself in varied more and in wider terms than the two previous cases, and are therefore presented in figure 6. A noticeable score increment in her raw scores can be observed at weeks 10 and 12.

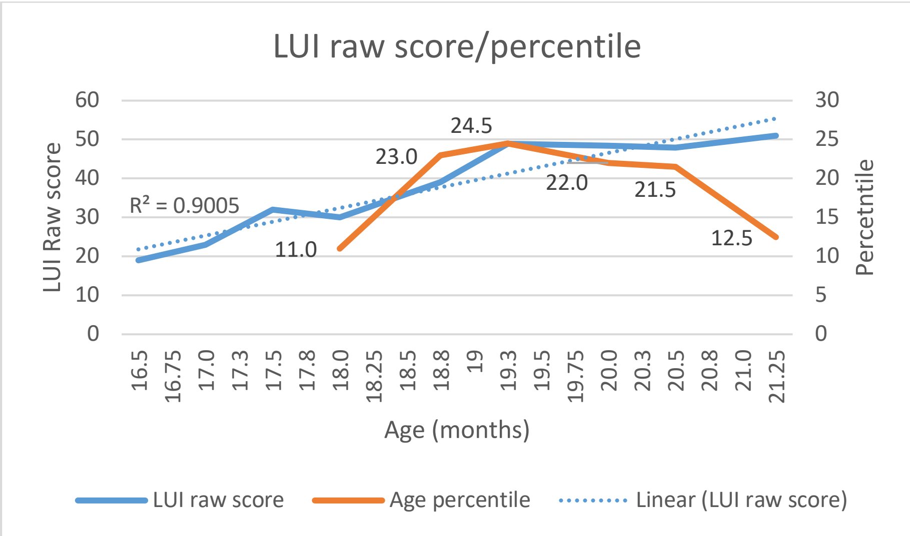


Figure 6. Daisy's LUI raw scores and age percentiles, per month of age. The dotted blue line represents a fitted linear regression model.

Question	Score
1.1 My child understands everything I say to they	1
1.2 My child understands sentences	1
1.3 Sometimes, my child understands me, but pretends otherwise	-3
1.4 Sometimes, my child does not cooperate because they do not understand what I'm asking for	2
1.5 My child has learned to walk	3
2 How autonomous would you say your child is?	4

Table 11. Initial questions concerning the mother's appraisal of Daisy's development and autonomy. Questions 1.1-5 correspond to Likert scales where -3 = I do not agree at all, and 3 = I completely agree. Question 2 corresponds to a Likert scale where 0 = Very dependent, and 10 = Very autonomous.

No particular changes in the prosody of Daisy's mother can be observed following the aforementioned weeks, thus providing no support for hypothesis 2.4.1. However, 12 prosodic measures evidence significant correlations. Remarkably, in Daisy's case significant correlations seem to consistently cluster around speech rate and f0-related measures (Table 12). In terms of the latter, the f0 standard deviation, maxima and range of her mother's IDSp and pedagogic IDSp seem to have decreased as Daisy's pragmatic competence augmented. Such correlations thus present for the first time more substantial support for hypothesis 2.4.2.

On the other hand, on the exception of gratifying IDSp, all of Daisy's mother's speech rate measures returned correlated negatively with LUI raw scores. Because parents' speech rate is expected to rather increase their children develop, this finding remains puzzling and does not support hypothesis 2.4.2.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	-0.2	0.0	0.0	0.3	0.2	0.3	-0.3	-0.3	-	0.3		
Ch	/	0.0	-0.2	-0.4	-0.3	-0.3	0.0	-0.1	-0.1	/		
ID	-0.7**	-0.4	-0.4	-0.6**	-0.5*	-0.6*	0.3	0.1	0.5*	-0.1	-0.4	0.3
Inq	-0.6**	-0.1	-0.1	0.0	-0.1	0.0	0.4	0.4	0.2	0.2		
Grat	0.1	0.3	0.3	0.0	0.3	0.2	0.3	0.4	0.1	-0.2		
Ped	-0.6**	-0.5*	-0.4	-0.5*	-0.6*	-0.6**	0.2	-0.1	0.3	-0.3		
Play	-0.6*	-0.2	-0.2	-0.2	-0.1	-0.1	0.3	0.1	0.3	-0.4		

* = $p < .05$, ** = $< p < .01$

Table 12. Pearson correlation coefficients between LUI raw scores and prosodic measures, per speaker, speech register and pragmatic function (for further details see previous cases). The cell corresponding to adult-directed f0 falls (Fls) is empty, given that Daisy's mother presented no value for that measure.

Average attentional bouts paid by Daisy to musical stimuli— both including and excluding interactional songs —are displayed in Figure 7. Similarly to Hugh's case, both fitted linear models are practically identical, and suggest that attentional bouts decreased as time progressed. This impression is confirmed by significant negative correlations between LUI raw scores and average attentional bouts, presented in Table 13. Also in a similar fashion to the first case, an increment can be observed around the introduction of interactional songs. Interactional songs seem do raise averages, as evidenced virtually every week after their introduction. No significant patterns from week to week were obtained when separated into the different genres (nursery rhymes/pop), degrees of familiarity (familiar/non familiar) or interaction (interactive/non interactive). Finally, as in Hugh's case, the fact that attentional bouts tend to decrease directly contradicts hypotheses 1.4.3 and 2.4.4.

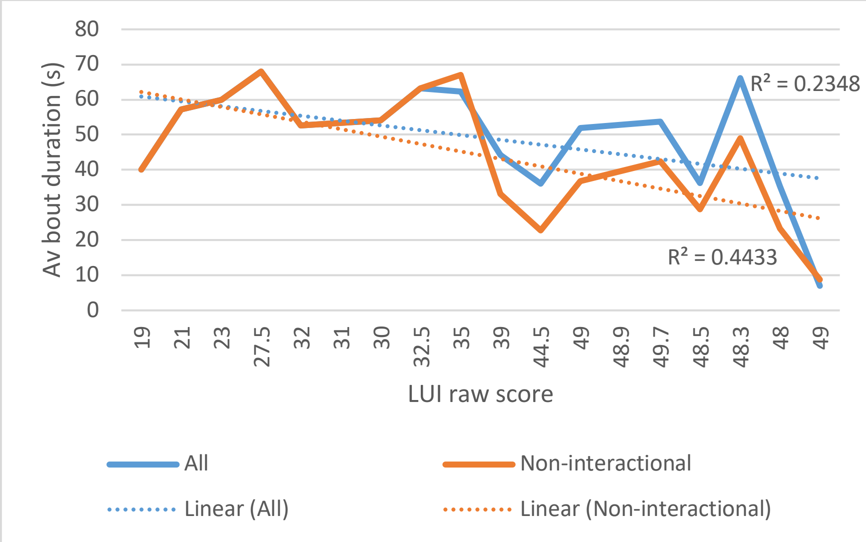


Figure 7. Lower left-hand side. Daisy's average attentional bout duration to all musical stimuli, in seconds (blue line), and average attentional bouts to all musical stimuli except those involving interaction (InterNurs and InterPop), in seconds (orange line), per LUI raw score. Dotted lines represent corresponding fitted linear regression models.

In terms of correlations between prosodic measures and average attentional bouts (excluding interactional songs), results are consistent with those reported in the previous table. As speech rates diminished, so did the length of attentional bouts (table 14). The number is significant correlations is 14. The same applies to the f0 mean, standard deviation, maximum and range of IDSp, the median, mean, standard deviation, maximum and range of pedagogic IDSp, and median, mean, maximum and range of playful IDSp.

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
Correlation	-0.4	-0.5	-0.7**	-0.8**	0.1	-0.4	-0.5*	-0.7**

* = p < .05, ** = < p < .01

Table 13. Pearson correlations coefficients between LUI raw scores and average attentional bouts, per genre, familiarity, and whether they were associated with interaction or not (see previous cases).

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	0.2	-0.2	-0.3	-0.6*	-0.5	-0.6	-0.2	-0.2	-	-0.1		
Ch	/	0.0		0.2	0.2	0.3	-0.1	-0.3	0.0	/		
ID	0.6**	0.4	0.5*	0.6**	0.7**	0.6**	-0.5*	-0.3	-0.6	0.1	0.2	0.2
Inq	0.5	0.3	0.4	0.3	0.2	0.2	-0.4	-0.3	-0.4	-0.1		
Grat	0.1	0.2	0.1	-0.1	-0.1	-0.1	-0.5*	-0.4	-0.3	0.0		
Ped	0.5*	0.5*	0.6**	0.7**	0.7**	0.7**	-0.4	-0.2	-0.4*	0.2		
Play	0.4	0.6*	0.5*	0.4	0.5*	0.4	-0.6*	-0.2	-0.5*	0.1		

* = p < .05, ** = < p < .01

Table 14. Pearson correlation coefficients between prosodic measures and average attentional bouts excluding interactional songs, per speaker, speech register and pragmatic function (see previous cases). The cell corresponding to adult-directed f0 falls (Fls) is empty, given that Daisy's mother presented no value for that measure.

3.1.3.2. Qualitative appraisal

Daisy is little girl who already talks and walks on her own by the start date of the study. She kept happy and calm in the vast majority of the situations she was exposed to, and her relationship with her mum seemed particularly harmonious, with little frustration-driven conflict. These impressions are corroborated by scores in Table 11. For instance, her mother's reported very little deceptive attitudes (question 1.3) and a fair sense of autonomy (question 2).

Most sessions took place in a quiet spare room in her house, and once in the guest room of a relative they were visiting. She was always at a short distance from her mum, and generally had one or two toys to play with autonomously (though she had none during in a couple of sessions). Such a setting was later described by her mum as pleasant, yet unnatural (Appendix J). Pleasant insofar they were together, unnatural because they would not normally sit in that room.

Quite from the outset and throughout the study, Daisy proved to be particularly able to listen and focus on music for rather long bouts, and with comparatively-less need for novel or parallel stimuli such as toys. From the third session onwards, her mum suggested that she climb on her lap facing outwards. From that moment onwards, Daisy mostly assumed such a position whenever music would start. The researcher noticed that such a dynamic resembled that of a toddlers music group. A post-hoc interview confirmed such an intuition, as both mother and daughter had been attending a music group by the time the study started. The combination between Daisy's experience in music groups and her mother's suggestion of her assuming the same position during the third session could partially explain the peak on average attentional bout duration observed on that date. Her musical experience could also explain the fact that, unlike the previous cases, the introduction of interactional music during week 8 did not generate a peak attention average (in Hugh and Louisa's cases it was the highest so far). It could be that this kind of interaction was not as novel and appealing to her. After week 10, Daisy started paying fewer attention to music— sometimes ignoring it completely—, and was more often distracted by toys or surrounding objects.

3.1.4. Tommy

3.1.4.1. LUI, vocal and attentional data

Boy, 23.8 months-old when started the study. His LUI raw scores show a slow but constant increase in his pragmatic skills, as corroborated by a Pearson's Linear Least Squares Regression (R^2 = .93). Unlike cases in the previous age group and as in the case of the next two participating children, Tommy's LUI raw scores could be compared to the rest of his age peers from the outset. He placed himself in the 1st percentile for most of the study's duration, to then show a steep increment in the last two questionnaires (Figure 8). Thus, abrupt changes in Tommy's pragmatic development are can be read in two ways. On the one hand, the LUI raw score rate of increment seems consistently steeper between weeks 1 and 3, from week 9 onwards. On the other hand, LUI percentile rate of increment seems consistently steeper from week 14 onwards.

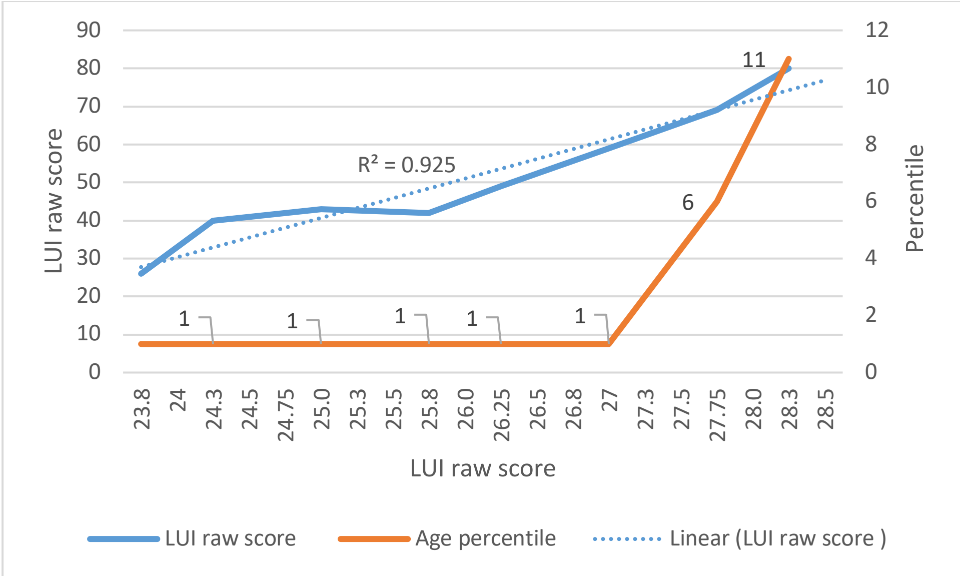


Figure 8. Tommy's LUI raw scores and age percentiles, per month of age. The dotted blue line represents a fitted linear regression model.

Question	Score
1.1 My child understands everything I say to they	1
1.2 My child understands sentences	1
1.3 Sometimes, my child understands me, but pretends otherwise	2
1.4 Sometimes, my child does not cooperate because they do not understand what I'm asking for	1
1.5 My child has learned to walk	3
2 How autonomous would you say your child is?	7

Table 15. Initial questions concerning the mother's appraisal of Tommy's development and autonomy. Questions 1.1-5 correspond to Likert scales where -3 = I do not agree at all, and 3 = I completely agree. Question 2 corresponds to a Likert scale where 0 = Very dependent, and 10 = Very autonomous.

In order to assess hypothesis 2.4.2, prosodic measures extracted weekly from recordings were correlated with their corresponding LUI raw scores. Most of the six significant correlations to be found seem to cluster around contour measures of IDSp and gratifying IDSp (Table 16). Some of these correlations behave as predicted by hypothesis 2.4.2 (e.g. negative correlations between LUI raw scores and IDSp f0 contour-related parameters) while others did not (e.g. positive correlations in gratifying IDSp). As a result, this case does not provide clear support for the mentioned hypothesis.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI	ID/AD	Ut/m
AD	0.1	-0.1	-0.1	-0.2	0.1	0.1	-0.3	/	-0.3	-0.1		
Ch	/	0.2	0.1	0.0	-0.1	-0.1	0.6	0.6	0.6	/		
ID	0.4	-0.1	0.0	0.0	0.1	0.2	-0.4*	-0.2*	-0.4*	0.4	-0.2	-0.4
Inq	0.2	-0.4	-0.4	-0.3	0.0	0.1	-0.2	0.0	-0.3	0.1		
Grat	0.5	0.6*	0.4	0.1	-0.2	0.0	0.6*	0.7*	-0.5	0.5		
Ped	-0.1	-0.4	-0.3	-0.1	-0.1	0.0	-0.3	-0.1	-0.3	0.4		
Play	0.0	0.4	0.3	0.1	0.2	0.2	-0.3	-0.1	-0.3	0.3		

* = p < .05, ** = < p < .01

Table 16. Pearson correlations coefficients between LUI raw scores and prosodic measures, per speaker, speech register and pragmatic function (see cases 1 or 2). The cell corresponding to adult-directed f0 rises (Rss) is empty, given that Tommy's mother presented no value for that measure.

Regarding hypothesis 2.4.1, a number of f0-related prosodic measures evidenced somewhat steep changes either following the first 3 weeks of the study, or the 10th. However, such changes are not maintained throughout the following weeks. Furthermore, they were not accompanied by significant correlations, and thus cannot be interpreted as support for the hypothesis. In Tommy's case, it seemed relevant to asses hypothesis 2.4.1 not just in terms of LUI raw scores, but also percentiles. Following week 14, the mother's IDSp f0 Mean evidenced a sudden dip in week 15 (Figure 11). However, such a dip in the f0 Mean does not seem to lead a general change in the mother's prosody, as it is followed by a continuous rise in the subsequent weeks. Furthermore, there was a positive, non-significant correlation between IDSp f0 Mean and LUI percentile.

Tommy's average attentional bouts elicited by musical stimuli— both including and excluding interactional songs —are displayed in Figure 9. Fitted linear models present different yet decreasing trends as time progressed. This is confirmed by a significant negative correlation between LUI raw scores and average attentional bouts elicited by non-interactional music stimuli (Table 17). As in the previous case, a peak in bout length cannot be observed around the introduction of interactional songs. Still, interactional songs do seem to raise averages in the following weeks, even more than markedly than in Daisy's case. No significant patterns from week to week were obtained when separated into the different genres (nursery rhymes/pop), degrees of familiarity (familiar/non familiar) or interaction (interactive/non interactive). Once again, the fact that attentional bouts tend to decrease directly contradicts hypotheses 1.4.3 and 2.4.4.

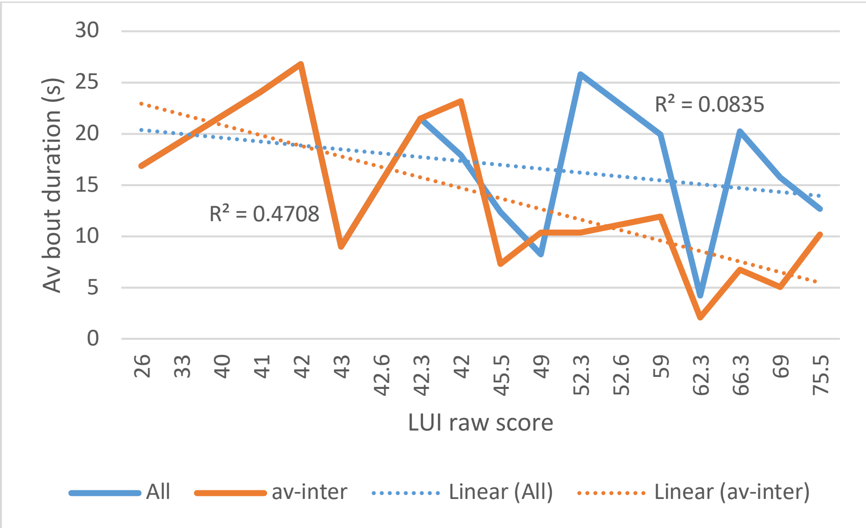


Figure 9. Lower left-hand side. Tommy's average attentional bout duration to all musical stimuli, in seconds (blue line), and average attentional bouts to all musical stimuli except those involving interaction (InterNurs and InterPop), in seconds (orange line), per LUI raw score. Dotted lines represent corresponding fitted linear regression models.

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
Correlation	-0.4	-0.5	-0.3	-0.6	0.5	-0.4	-0.3	-0.6*

* = p < .05, ** = < p < .01

Table 17. Pearson correlations coefficients between LUI raw scores and average attentional bouts, per genre, familiarity, and whether they were associated with interaction or not (see cases 1 or 2).

In Tommy's case, correlations between prosodic measures and average attentional bouts (excluding interactional songs), present more significant results than correlations in Table 16. Inquiring and pedagogic IDSp f0-related measures present positive correlations, confirming that they decreased as the average length of attentional bouts did (Table 18). Some IDSp f0 contour-related measures are significant, as in Table 16, this time accompanied by some A-D f0 contour-related measures.

	Prosodic measure												
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m	%Ped
AD	-0.3	0.4	0.4	0.5	0.3	0.3	0.8**	-	0.8**	0.0			
Ch	/	-0.2	-0.2	-0.4	-0.4	-0.3	-0.6*	0.5	-0.7**	/			
ID	-0.3	0.4	0.4	0.8**	0.8**	0.9**	0.5*	0.5*	0.3	0.1	0.2	0.3	
Inq	-0.3	0.8**	0.8**	0.8**	0.7**	0.5*	0.5*	0.3	0.5	0.4			
Grat	-0.5	0.0	-0.1	0.0	0.3	0.1	-0.1	0.0	-0.1	-0.1			
Ped	0.3	0.6*	0.6*	0.4	0.5	0.5	0.4	0.5*	0.2	0.1			0.5
Play	0.0	0.1	0.3	0.1	-0.1	-0.1	-0.2	0.2	-0.4	0.0			

* = p < .05, ** = < p < .01

Table 18. Pearson correlations coefficients between prosodic measures and average attentional bouts excluding interactional songs, per speaker, speech register and pragmatic function (see cases 1 or 2). The cell corresponding to adult-directed f0 rises (Rss) is empty, given that Tommy's mother presented no value for that measure.

3.1.4.2. Qualitative appraisal

Tommy is a lively boy full of energy, that often ran around in the room before his father could calm him down and get him to play on the floor. He seemed witty, at times somewhat naughty. Such an impression is coherent with his mother's scoring in question 1.3 in Table 15, where she conveys the notion that her son would rather often present uncooperativeness by means of deception. Also coherent is the fact that his mother perceived him as fairly autonomous boy (score of 7 in question 2), probably more used to actively manipulating the world around him according to his interests.

He could talk in monosyllables by the start date of the study. All sessions took place in his living room, with eventually him playing autonomously at a certain distance from his dad. He was capable of staying still and casually listening to music, sometimes autonomously. In a similar fashion to participant one (Hugh), he often displayed short, rather stereotypical responses to music, which he sometimes repeatedly alternated with short bouts of playing with his toys. Tommy had, more than once, stimuli that competed with music during the sessions. For instance, he once had his bottle of milk with him and seemed quite to focus on it. His dad sometimes distracted him by drawing his attention (against instructions) towards a toy. For these reasons, it can be speculated that in the case of this child, attentional bouts would have often been longer in other circumstances. Nevertheless, no stimulus could distract him from doing the actions associated with some of the nurseries. To the sound of 'The wheels on the bus', in particular, Tommy evidenced interest and concentration that frankly contrasted with that prompted by the rest of the repertoire he was exposed to. He would pay long bouts of sustained attention and assume quite a focused posture, and often repeat some of the monosyllabic keywords in the lyrics (e.g. ‘all’ for the sentence ‘all through the town’). He would also carefully perform each of the verses’ associated actions and prevent his dad from doing so whenever he gave them a try as if it was his exclusive right.

3.1.5. Patricia

3.1.5.1. LUI, vocal and attentional data

Girl, 23.4 months-old when started the study. Her LUI raw scores evidence, as expected, a slow but constant increase in her pragmatic skills, as corroborated by a Pearson’s Linear Least Squares Regression (R^2 = .87) (Figure 10). No noticeable raw score increases can be observed, rendering the assessment of hypothesis 2.4.1 and 2.4.4 impossible. Patricia systematically positioned herself in the 4th and 5th percentiles throughout the study.

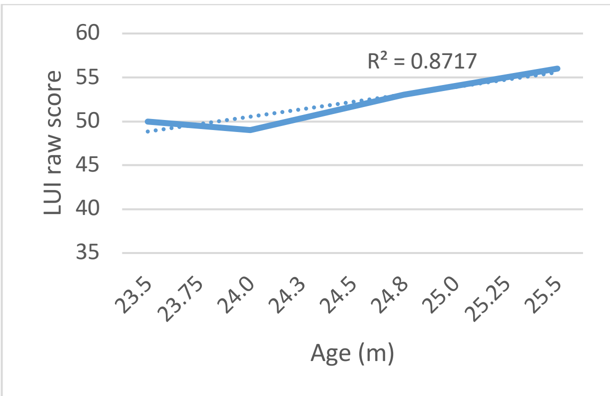


Figure 10. Patricia’s LUI raw scores per week of age.
The dotted line represents the R^2 fitted line.

Virtually no significant correlations were found when assessing the weekly evolution of her mother’s prosodic parameters (Table 19), thus providing no support for hypothesis 2.4.2. As presented In Figure 10, Patricia’s average attentional bout duration directed at recorded music shows little variation in terms of average length of attentional bouts to non-interactional recorded music (orange dotted line in the graph). The trend turns into a slight overall increment when also considering interactional songs (blue dotted line). No significant patterns from week to week were obtained when separated into the different genres (nursery rhymes/pop), degrees of familiarity (familiar/non familiar) or interaction (interactive/non interactive).

Question	Score
1.1 My child understands everything I say to they	3
1.2 My child understands sentences	3
1.3 Sometimes, my child understands me, but pretends otherwise	3
1.4 Sometimes, my child does not cooperate because they do not understand what I'm asking for	-2
1.5 My child has learned to walk	3
2 How autonomous would you say your child is?	9

Table 19. Initial questions concerning the mother's appraisal of Patricia's development and autonomy. Questions 1.1-5 correspond to Likert scales where -3 = I do not agree at all, and 3 = I completely agree. Question 2 corresponds to a Likert scale where 0 = Very dependent, and 10 = Very autonomous.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	0.1	0.2	0.2	0.5	0.1	-0.1	/	/	/	0.2		
Ch	/	0.2	0.1	-0.3	-0.3	-0.2	-0.2	0.5	-0.2	/		
ID	0.0	0.2	0.2	0.1	0.1	0.2	-0.4*	0.0	0.0	0.0	-0.4	-0.2
Inq	-0.1	0.2	-0.2	0.4	-0.1	-0.2	-0.4	-0.1	-0.3	-0.1		
Grat	-0.1	-0.1*	-0.3	0.0	-0.2	-0.2	/	/	-0.4	0.0		
Ped	-0.1	0.1	0.4	0.2	0.2	0.2	0.0	-0.2	-0.1	-0.3		
Play	-0.1	0.0	0.2	0.2	0.0	0.0	0.3	0.0	/	-0.1		

* = p < .05, ** = < p < .01

Table 20. Pearson correlations coefficients between LUI raw scores and prosodic measures, per speaker, speech register and pragmatic function (see cases 1 or 2). A '/' is placed whenever Patricia's mother presented no value for a given measure.

In agreement with the rather horizontal trendlines in Figure 10, only small, non-significant correlations between Patricia's LUI raw scores and average attentional bouts (Table 21), and between her mother's prosodic measures and her own attentional bouts (Table 22) can be found. These finding thus provide no support for hypotheses 2.4.3 and 2.3.4.

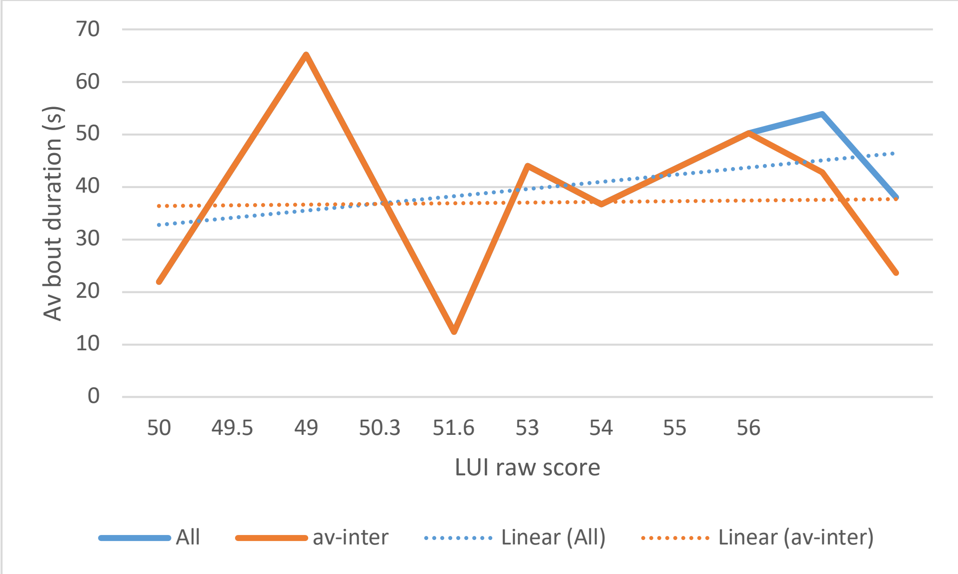


Figure 10. Patricia's average attentional bout duration to all musical stimuli, in seconds (blue line), and average attentional bouts to all musical stimuli except those involving interaction (InterNurs and InterPop), in seconds (orange line), per LUI raw score. Dotted lines represent corresponding fitted linear regression models.

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
Correlation	-0.2	-0.1	-1.0	-0.2	/	/	0.0	0.0

* = p < .05, ** = < p < .01

Table 21. Pearson correlations coefficients between LUI raw scores and average attentional bouts, per genre[26], familiarity, and whether they were associated with interaction or not (see cases 1 or 2).

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	0.1	0.2	0.2	0.5	0.1	-0.1	/	/	/	0.2		
Ch	/	0.2	0.1	-0.3	-0.3	-0.2	-0.2	0.5	-0.2	/		
ID	0.0	0.2	0.2	0.1	0.1	0.2	-0.4	0.0	0.0	0.0	-0.4	0.2
Inq	-0.1	0.2	-0.2	0.4	-0.1	-0.2	-0.4	-0.1	-0.3	-0.1		
Grat	-0.1	-0.1	-0.3	0.0	-0.2	-0.2	/	/	-0.4	0.0		
Ped	-0.1	0.1	0.4	0.2	0.2	0.2	0.0	-0.2	-0.1	-0.3		
Play	-0.1	0.0	0.2	0.2	0.0	0.0	0.3	0.0	/	-0.1		

* = p < .05, ** = < p < .01

Table 22. Pearson correlations coefficients between prosodic measures and average attentional bouts excluding interactional songs, per speaker, speech register and pragmatic function (see cases 1 or 2). The cell corresponding to adult-directed f0 rises (Rss) is empty, given that Tommy's mother presented no value for that measure.

[26] As explained in section 2.1.1 in this chapter, Patricia participated for 10 weeks in the study. Therefore, InterNurs and InterPop where presented at the end of her participation, and do not provide enough datapoints for correlations to be drawn.

3.1.6.2. Qualitative appraisal

Patricia is a little girl with a somewhat serious look and an engaging attitude. She was already able to walk by the study's start date. Al sessions took place in her living room and started with her being called by her mum. At the beginning of the first two sessions— against the researcher's instructions —her mother conveyed the idea that they were about to listen to some music. The mother was thus advised against it, and such a behaviour was not repeated from the third session onwards. When considering statements from a post-hoc interview, it seems likely that Patricia may have felt somewhat disoriented in the proposed setting.

Patricia's behaviour immediately detached in two aspects from the rest of the participating children so far reported. On the one hand and from the outset, she payed long attentional bouts to nursery rhymes and play songs. She would stare at the device where music was coming from and more often than not listen until the very end of the piece. She would sometimes perform actions, dance and smile, and direct these behaviours to her mum. In a number of other instances, she would show great focus and a rather serious face, as if she was trying to understand or decipher something. On the other hand, her reaction to all pop songs presented to her was diametrically opposed. After a few seconds of playback Patricia would, almost invariably, show signs of annoyance and emphatically demand for the music to stop. In one of the sessions the vehemence of her reaction translated into her shouting 'no!' while slapping the smartphone (the source of music) out of her mum's hand and sending it flying. At the same time, her mother assigned the highest possible score (3 out of 3) to question 1.3 in Table 19. Such a score conveys the notion that Patricia would often present uncooperativeness by means of deception. Additionally, her mother perceived her as very autonomous (9 out of 10 in question 2). In the same vein, the post-hoc interview revealed that Patricia was going through a 'stubborn streak' for some months already, an attitude that seems to have been somewhat intensified by the setting. Taken as a whole, data depicted a little girl that seems to be able to quickly identify what she likes or is interested in and a has the capacity focus on/enjoy it. At the same time, she seemed to be perfectly able to identify what she did not like or was not interested in and evidenced a will to reject it.

3.1.6. Gwendolyn

3.1.6.1. LUI, vocal and attentional data

Girl, 22 months-old when started the study. Her LUI raw scores show, as expected, a slow but constant increase in her pragmatic skills, as corroborated by a Pearson's Linear Least Squares Regression (R^2 = .94). The percentiles she located herself in increased rather steadily starting 34th and ending at the 73rd (Figure 11). No noticeable raw score increases can be observed, making the assessment of hypothesis 2.4.1 and 2.4.2 impossible.

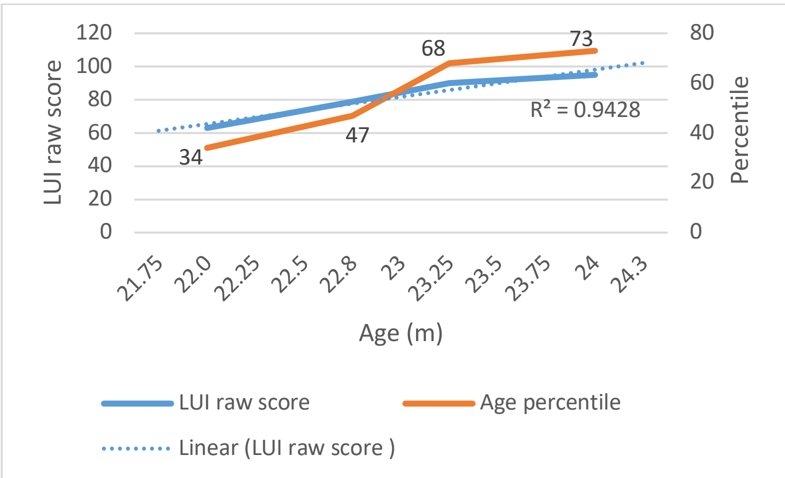


Figure 11. Gwendolyn's LUI raw scores and age percentiles, per month of age. The dotted blue line represents a fitted linear regression model.

Question	Score
1.1 My child understands everything I say to they	2
1.2 My child understands sentences	3
1.3 Sometimes, my child understands me, but pretends otherwise	-3
1.4 Sometimes, my child does not cooperate because they do not understand what I'm asking for	-3
1.5 My child has learned to walk	3
2 How autonomous would you say your child is?	7

Table 23. Initial questions concerning the mother's appraisal of Gwendolyn's development and autonomy. Questions 1.1-5 correspond to Likert scales where -3 = I do not agree at all, and 3 = I completely agree. Question 2 corresponds to a Likert scale where 0 = Very dependent, and 10 = Very autonomous.

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI _	ID/AD	Ut/m
AD	0.1	0.0	0.1	-0.1	0.1	0.0	-0.4	-0.2	-0.3	0.2		
Ch	/	0.4	0.4	0.5	0.4	0.4	-0.1	0.2	-0.1	/		
ID	0.0	0.1*	0.3	0.4	0.4	0.4	0.5*	0.3	0.3	0.0	0.1	0.4
Inq	/	/	/	/	/	/	/	/	/	/		
Grat	-0.1	0.0	0.3	0.3	0.4	0.4	-0.3	-0.4	0.4	0.5		
Ped	0.1	0.2	0.2	0.4	0.4	0.4	0.4	0.4	0.2	0.1		
Play	0.1	0.1	0.1	0.3	0.3	0.4	-0.1	0.4	-0.2	0.2		

* = p < .05, ** = < p < .01

Table 24. Pearson correlations coefficients between LUI raw scores and prosodic measures, per speaker, speech register and pragmatic function (see cases 1 or 2). The cell corresponding to adult-directed f0 rises (Rss) is empty, given that Tommy's mother presented no value for that measure.

As in the previous case, virtually no significant correlations were found when assessing the weekly evolution of her mother's prosodic parameters (Table 24). Aggregated IDSp constituted an exception, with a weak correlation with f0 median and a moderate one with the number of glissandi. In sum, evidence in this case does not provide support for hypothesis 2.4.2.

As shown in Figure 12, Gwendolyn's average attentional bout duration directed at recorded music shows a rather steady development throughout the weeks of the study, until the last two sessions— corresponding to the introduction of interactional songs —, where it peaks. No significant patterns from week to week were obtained when separated into the different genres (nursery rhymes/pop), degrees of familiarity (familiar/non familiar) or interaction (interactive/non interactive). Similar to Patricia's case, only small and moderate non-significant correlations between Gwendolyn's LUI raw scores and average attentional bouts (Table 25) can be found. Something similar can be reported concerning correlations between her mother's prosodic measures and her own attentional bouts (Table 22) can be found. By this token, once again evidence does not provide support for hypothesis 2.4.3 and 2.4.4.

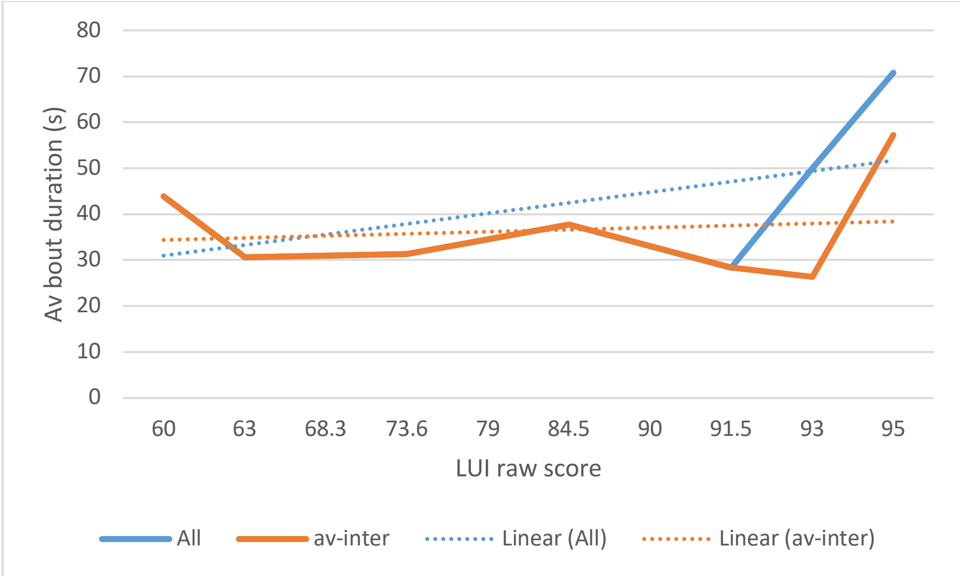


Figure 12. Lower left-hand side. Gwendolyn's average attentional bout duration to all musical stimuli, in seconds (blue line), and average attentional bouts to all musical stimuli except those involving interaction (InterNurs and InterPop), in seconds (orange line), per LUI raw score. Dotted lines represent corresponding fitted linear regression models.

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
Correlation	-0.2	-0.1	0.3	0.4	/	/	0.4	0.1

* = p < .05, ** = < p < .01

Table 25. Pearson correlations coefficients between LUI raw scores and average attentional bouts, per genre[27], familiarity, and whether they were associated with interaction or not (see cases 1 or 2).

	Prosodic measure											
	SpR	Med	Mn	SD	Mx	Rg	Gls	Rss	Fls	nPVI_	ID/AD	Ut/m
AD	-0.3	-0.1	-0.2	-0.2	-0.3	-0.2	0.7	0.5	0.6	0.5		
Ch	/	-0.4	-0.4	0.0	-0.4	-0.3	-0.2	0.6	-0.3	/		
ID	-0.1	-0.3	-0.2	-0.3	-0.3	-0.4	-0.8*	-0.2	-0.7	0.0	-0.3	-0.6
Inq	/	/	/	/	/	/	/	/	/	/		
Grat	0.5	0.5	0.3	0.2	0.1	0.1	0.6	0.5	-0.5	0.2		
Ped	-0.1	-0.1*	0.1	0.1	-0.1	-0.2	-0.5	-0.1	-0.6	-0.3		
Play	-0.3	-0.5	-0.6	-0.4	-0.6	-0.5		0.1		-0.6		

* = p < .05, ** = < p < .01

Table 26. Pearson correlations coefficients between prosodic measures and average attentional bouts excluding interactional songs, per speaker, speech register and pragmatic function (see cases 1 or 2). The cell corresponding to adult-directed f0 rises (Rss) is empty, given that Tommy's mother presented no value for that measure.

[27] Regarding the lack of data for InterNurs and InterPop, see explanation in the previous case's equivalent table.

3.1.6.2. Qualitative appraisal

Gwendolyn is little girl that also gave an impression of seriousness, and a somewhat inquisitive stance. Like the previous participating child, she was already able to walk by the study's start date. In her case, sessions took place in different rooms of her house, or relatives they visited. In any case, her and her mum were always alone and undisturbed.

Gwendolyn's response to the different kinds of music presented to her had remarkable similarities with that of the previous case (Patricia). Also from the outset, she paid long attentional bouts to nursery rhymes and play songs. She would focus her gaze on the device and stay still, listening attentively. She would sometimes perform actions related to the songs, and more often than not listen until the very end of the piece. Gwendolyn's reaction to pop songs was slightly less vehement than that of Patricia. One notable exception to this concerned the pop song her mum had selected as preferred piece for interaction. During the study, her mother reported her daughter repeatedly asking her to stop the music after a few seconds of the introduction. Still, such a rejection did not involve any impulsive physical components such as the one evidenced in the previous participant. It is interesting to notice that Gwendolyn mother did not judge her being as autonomous as Patricia (Table 23, question 2), and completely discarded intents of uncooperativeness or deception, as assessed by question 1.3.

3.2. Overall data and between-groups differences

3.2.1. Prosodic data

In this section, data will no longer be presented on weekly basis, but aggregated first by participant, then by group and finally collapsed as a whole. Figure 13a shows the value of several f0-related measures extracted from the average of all of Hugh and his mother's utterances throughout the weeks of the study. The mother's utterances have been separated in terms of register: ADSp or IDSp. IDSp was additionally sorted by pragmatic function (Gratifying, Pedagogic, Inquiring and Playful utterances). The same can be seen for Louisa and Daisy in Figures 13b and 13c respectively.

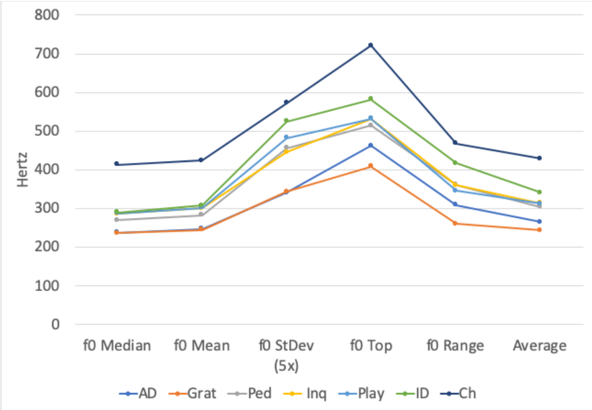


Figure 13a. Hugh's mother's average f0-related measures per speaker, register and function

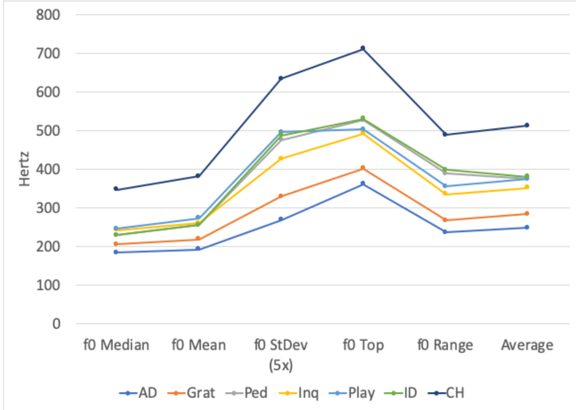


Figure 13b. Louisa's mother's average f0-related measures per speaker, register and function

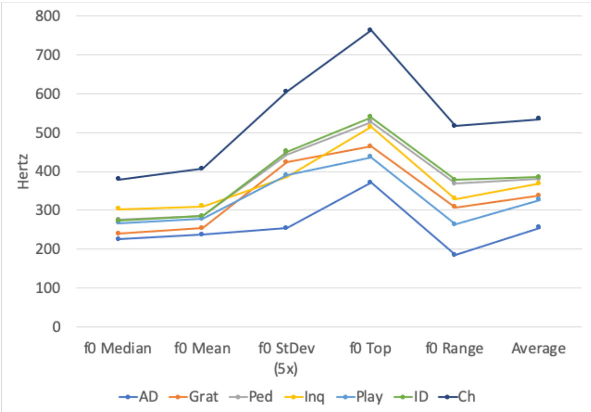


Figure 13c. Daisy's mother's average f0-related measures per speaker, register and function

Figures 13a-c show the average f0-related measures in the speech of the mothers of the group of 15-month-old children. The blue line (AD) represents the average of all the mothers' ADSp utterances. Similarly, orange (Grat) represents her Gratifying IDSp utterances, grey (Ped) represents Pedagogic IDSp, yellow (Inq) represents Inquiring IDSp, sky-blue (Play) represents Playful IDSp, and green (ID) represents IDSp without distinction of pragmatic function. Finally, the deep blue line (Ch) represents all of the Child's vocalizations. Displayed f0-related measures correspond to f0 Median, Mean, Maximum (Top) and Range. The f0 standard deviation has been multiplied 5x in order to allow for visual comparison. The average of the above is also presented.

It can be observed that in all three cases the Child's f0 parameters and their average are higher than those of the mother's without exception, their f0 Maxima between 700 and 800 hertz. With the exception of two of Hugh's mother's parameters— (f0 Maximum and Range) that match Gratifying IDSp —, ADSp utterances consistently show the lowest f0-related measures and f0 parameters average. In terms of pragmatic functions, Playful utterances appear as the highest in overall f0 parameters, and Gratifying one of the lowest. IDSp utterances (i.e. the totality of IDSp utterances recorded including further pragmatic functions than the four aforementioned) show higher f0 parameters than any of the individual, disaggregated pragmatic functions.

An equivalent analysis on the 23-month-old cases returns some similar and some different results to the previous group (Figures 14a-c). As the main point of similarity, and with the exception of two of Tommy's mother's parameters (again, f0 Maximum and Range), ADSp remains largely distinct from the rest of the vocal data, with the lowest f0.

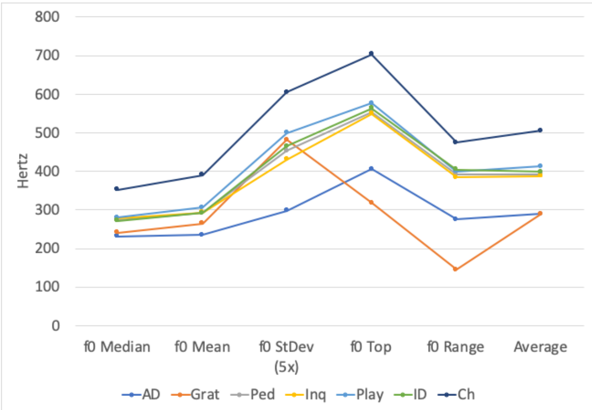


Figure 14a. Tommy's mother's f0-related measures per speaker, register and function

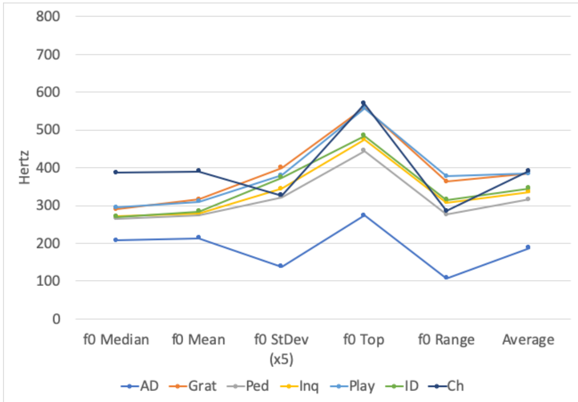


Figure 14b. Patricia's mother's f0-related measures per speaker, register and function

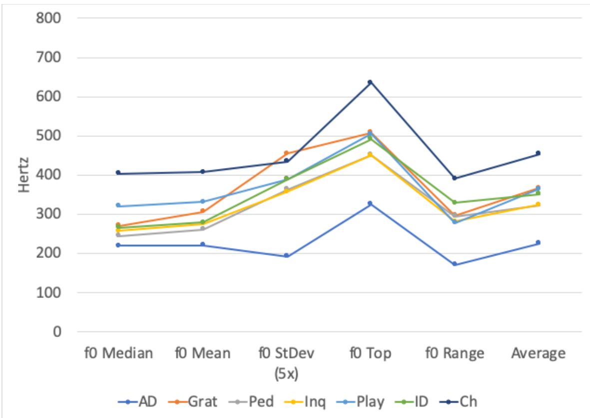


Figure 14c. Gwendolyn's mother's f0-related measures per speaker, register and function

Figures 14 a, b and c show the average f0-related measures in the speech of the mothers of the group of 23-month-old children. The blue line (AD) represents the average of all the mothers' ADSp utterances. Similarly, orange (Grat) represents her Gratifying IDSp utterances, grey (Ped) represents Pedagogic IDSp, yellow (Inq) represents Inquiring IDSp, sky-blue (Play) represents Playful IDSp, and green (ID) represents IDSp without distinction of pragmatic function. Finally, the deep blue line (Ch) represents all of the Child's vocalizations. Displayed F0-related measures correspond to f0 Median, Mean, Maximum (Top) and Range. The f0 standard deviation has been multiplied 5x in order to allow for visual comparison. The average of the above is also presented.

One primary difference is that the different IDSp pragmatic functions appear slightly more clustered together than in the previous group. Such an impression is corroborated by comparing the average standard deviation of all of Patricia and Gwendolyn's f0 parameters (72Hz) to that of the four other cases (75Hz). Another difference seems to be that Patricia and

Gwendolyn's f0 parameters are not as clearly detached from their mothers as the rest of the cases. These two girls' f0 Maxima is around 600 hertz, thus 100 lower than the other cases.

One notable exception to this was the average percentage of weekly pedagogic IDSp utterances out of the total weekly IDSp utterances (including gratifying, playful, etc.). As displayed in table 27, all three children in the first group received lower numbers of pedagogic IDSp utterances than any children in the second one. Accordingly, the group averages are significantly different, ($t(45) = -6.080$, $p < .001$). Furthermore, there is a significant, correlation between the two variables ($r(4) = 0.794$, $p = .05$)

Tables 4, 8, and other equivalents reported each participant's number of significant correlations between LUI raw scores and maternal prosody. When assessing these correlations beyond the individual level, a trend can be noticed: the number of significant correlations decreases with age (Table 27). Although the relationship between these two variables is not quite linear ($r = -0.787$, $p = .06$), both groups do differ significantly in terms of the mentioned correlations ($t(4) = 3.324$, $p = .03$), with the younger group having a number three times higher than its counterpart.

	P1	P2	P3	P4	P5	P6	Av
IDSp Ped %	25	34	60	66	81	61	40
Group av %		40			69		
# of sig. corr.	11	7	12	6	2	1	40
Group av %		10			3		

Table 27. First two rows: average percentage of weekly pedagogic IDSp utterances out of the total weekly IDSp utterances, per participant and group. Third and fourth rows: number of significant correlations between LUI raw score and maternal prosodic measures, per participant and group.

When collapsing all data into one single set, general prosodic profiles were distilled for each f0-related measure. As presented in Figure 15, speakers and registers detach from each other in terms of every assessed measure, with children vocalizing at higher frequencies than their mothers. In terms of prosodic exaggeration, ADSp can be observed being distinctly less exaggerated than IDSp in terms of every measure, both in its aggregated form or separated into pragmatic functions. The latter are less separated one to another, and their relative rank of exaggeration depends on the measure observed (e.g. playful IDSp has the highest f0 median and mean when compared to the other pragmatic functions and aggregated IDSp, but the same is not true when it comes to f0 standard deviations or maxima).

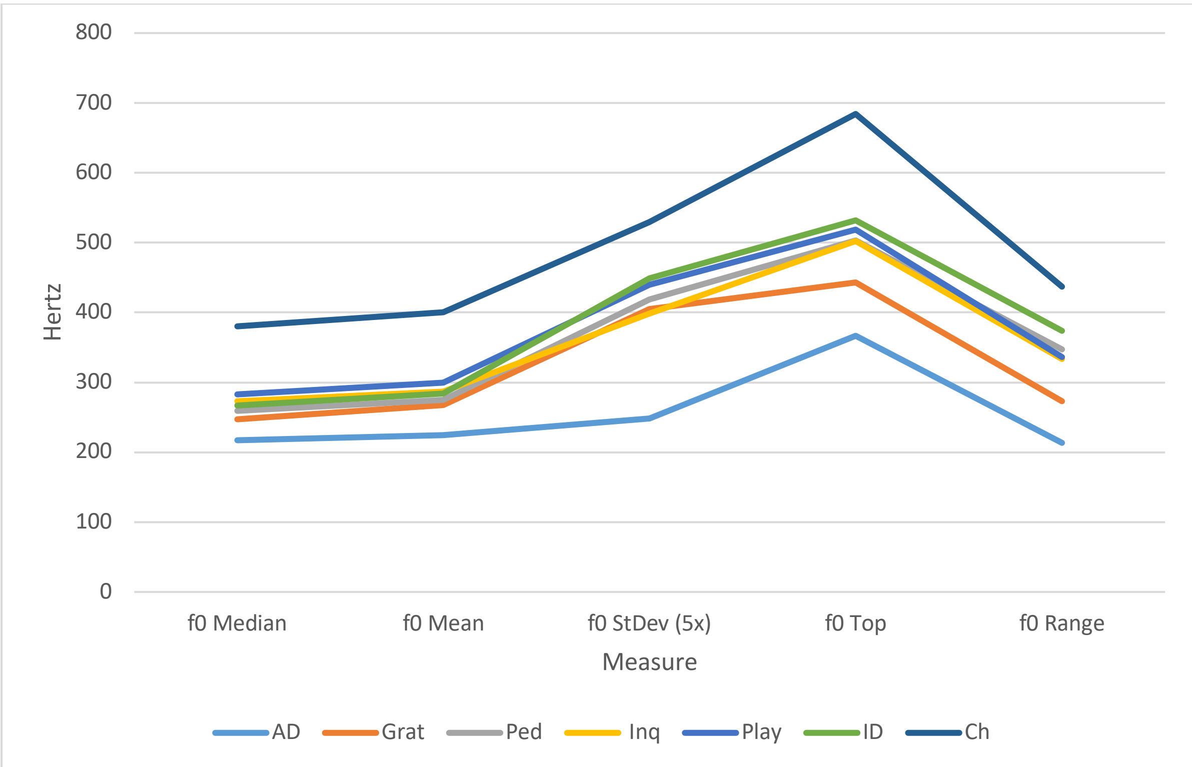


Figure 15. Prosodic profiles. The sky-blue line (AD) represents the average of all the 6 mothers' ADSp utterances. Similarly, orange (Grat) represents their averaged gratifying IDSp utterances, grey (Ped) represents pedagogic IDSp, yellow (Inq) represents inquiring IDSp, blue (Play) represents Playful IDSp, and green (ID) represents IDSp without distinction of pragmatic function. Finally, the deep blue line (Ch) represents all of the Child's vocalizations. F0-related measures shown are f0 Median, Mean, Maximum (Top) and Range. The f0 standard deviation has been multiplied 5x in order to allow for visual comparison.

As a further result of collapsing all data into one single set, only a few parameters returned relevant findings. Interestingly, IDSp nPVI was half of its ADSp counterpart (Figure 16) and the same pattern could be found regarding f0 contour glissandi, rises and falls (Figure 17). Inquiring and (especially) Gratifying IDSp utterances presented larger numbers of f0 contour Glissandi. Gratifying IDSp also had a comparatively larger number of f0 contour falls.

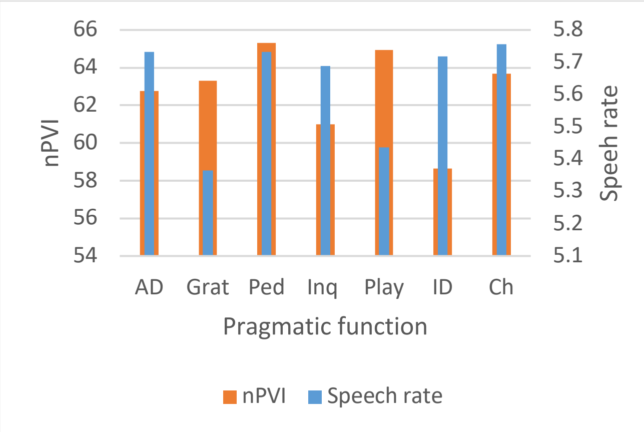


Figure 16. Speech Rate, and Syllable Nuclei Duration nPVI per speaker (child [Ch] or mother), register (adult-directed [AD] or infant-directed), and infant-directed pragmatic function (gratifying [Grat], pedagogic [Ped], Inquiring [Inq] and playful [Play]). Column height represents the average of all 6 cases.

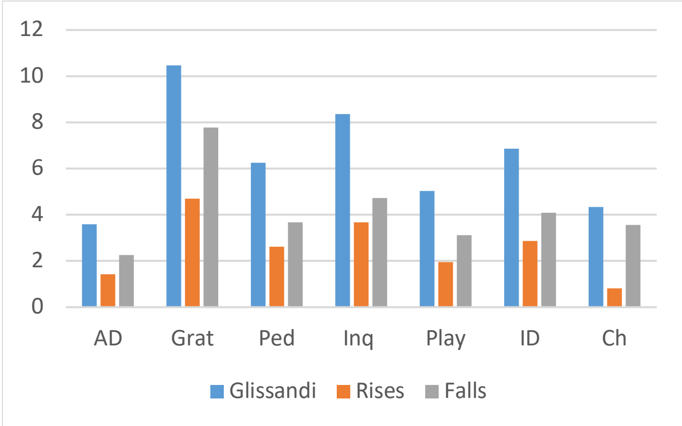


Figure 17. Number of glissandi, rises and falls per speaker (child [Ch] or mother), register (adult-directed [AD] or infant-directed), and infant-directed pragmatic function (gratifying [Grat], pedagogic [Ped], Inquiring [Inq] and playful [Play]). Figures represent the average of all 6 cases. Speech Rate figures have been multiplied 10x in order to allow visual comparison.

Table 28 shows the average attentional bout duration that each child paid to the different types of music presented to them. Attentional bout durations ranged from 1 second (a very short, initial fixation paid to any salient stimulus) up to over two minutes (the full length of a song). The average attentional bout duration of both groups is practically the same: 29 seconds for the 15-month-old and 31 for the 23-month-old. However, three girls (Daisy, Patricia and Gwendolyn) had average attentional bout durations that doubled those of the other three children. Hugh, Louisa and Tommy had average attentional bouts of 17, 21 and 17s, thus corresponding to the definition of cumulative fixation as defined in chapter 5, section 4.1. On the other hand, Daisy, Patricia and Gwendolyn's equivalents (49, 41 and 42s respectively) exceed initial fixation by a wide margin and can therefore safely be considered as sustained attention. The relative capacity of the different music types to elicit sustained attention was virtually the same across groups (Figure 19c), the main difference being that

23-month-olds pay even less attention to pop music, be it familiar or not and with or without interaction.

As displayed in Figure 18, nursery rhymes through which parents had previously interacted with their children (Inter Nurs) elicited, in average, the longest attentional bouts. The average attentional bout duration elicited by nursery rhymes through which participating mothers had previously interacted with children in the 23-month-old group doubles the equivalent in the 15-month-old group (Table 28). Such a difference is mainly explained again by Patricia and Gwendolyn's attentional bouts directed at such type of music: 133 and 122 seconds respectively. Second and third longest average attentional bouts were elicited by familiar and unfamiliar nurseries.

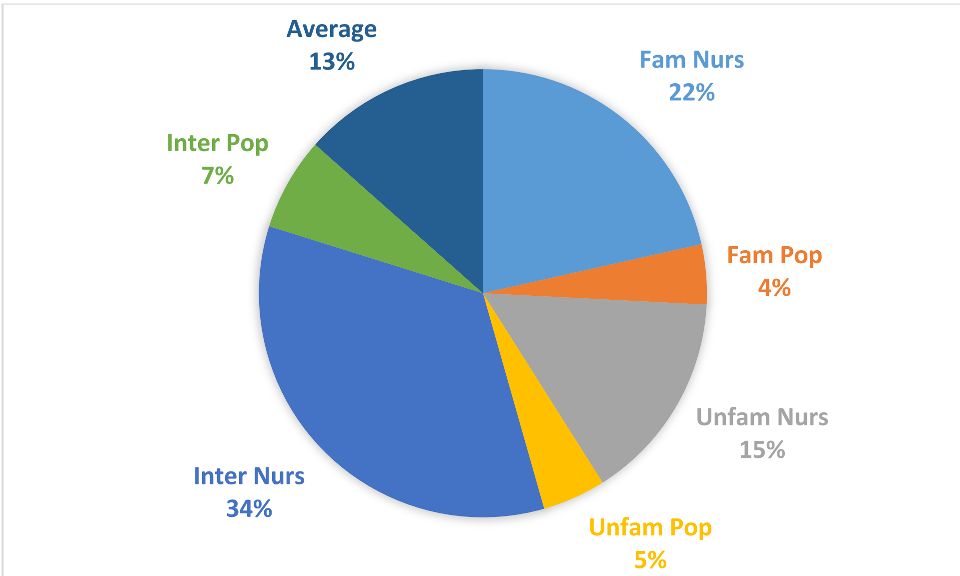


Figure 18. Comparative percentage of average attention paid by the children to the different music types presented to them. The latter comprise familiar nursery rhymes (FamNurs), familiar pop songs (FamPop), unfamiliar nursery rhymes (UnfamNurs), unfamiliar pop songs (UnfamPop), nursery rhymes through which parents had previously interacted with their children (InterNurs), and pop songs through which parents had previously interacted with their children (Inter pop), and the overall average.

	Participant	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Average
15-m-o Group	P1	27	8	17	6	31	0	17
	P2	44	10	10	12	40	32	21
	P3	72	27	42	24	97	21	49
	Average	48	15	23	14	56	18	29
	Participant	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Average
23-m-o Group	P4	24	5	11	4	49	4	17
	P5	67	1	67	5	133	20	41
	P6	62	9	64	12	123	Rejected	42
	Average	51	5	47	7	101	12	33
	Grand average	49	10	35	11	79	15	31

Table 28. Average attention (in seconds) paid to music per participant and music type. The latter comprise familiar nursery rhymes (FamNurs), familiar pop songs (FamPop), unfamiliar nursery rhymes (UnfamNurs), unfamiliar pop songs (UnfamPop), nursery rhymes through which parents had previously interacted with their children (InterNurs), and pop songs through which parents had previously interacted with their children (InterPop).

As previously mentioned, nursery rhymes through which mothers had previously interacted with their children (Inter Nurs) elicited, on average, the longest attentional bouts. Such attentional bouts were significantly larger than the overall average when compared through a one-tailed Student-t test ($t(5) = 2.496$, $p = .02$). Similar is the case of familiar nursery rhymes, which stood roughly above the overall average ($t(5) = 1.76$, $p = .054$). Two amid the presented forms of pop songs were significantly below the average: familiar pop songs ($t(5) = -3.06$, $p < .01$) and unfamiliar pop songs ($t(5) = -3.11$, $p < .01$). Pop songs through which mothers had previously interacted with their children were also below the overall average, but not at a statistically-significant level ($t(5) = -1.52$, $p = .08$). In order to correct for multiple comparisons, a one-way ANOVA was performed. Results confirmed a significant effect of music type on average attentional bout length [$F(5, 35) = 7.74$, $p < .001$].

Participants in the 15-month-old group presented virtually no active rejection to any of the music that was presented to them thorough the weeks (Table 29). On the other hand, Patricia and Gwendolyn in the 23-month-old group presented averages of 1.1 and 1.6 respectively, doubling and tripling the overall average. Student-t tests confirmed that these averages significantly detached from the overall average ($t(10) = 1.88$, $p = .04$; $t(10) = 4.5$, $p < .001$). Once again, in order to correct for multiple comparisons, a one-way ANOVA was performed. Results confirmed a significant effect of music type on average attentional bout length [F(5, 35) = 7.16, p < .001]. The few rejections that participants 1-4 showed occurred during the last 5 weeks of the study, whereas Patricia and Gwendolyn's rejections occurred from week 3 and 1 onwards, respectively. Furthermore, Gwendolyn consistently rejected Interactive Pop, to the point that it had to be discarded from the study.

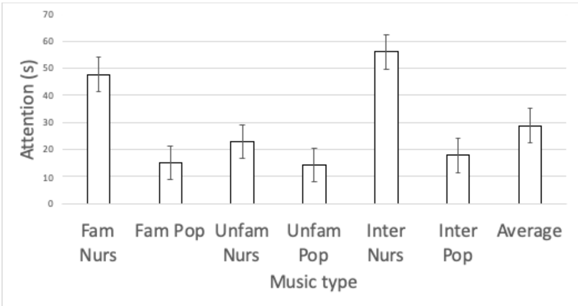


Figure 19a. 15-month-old group

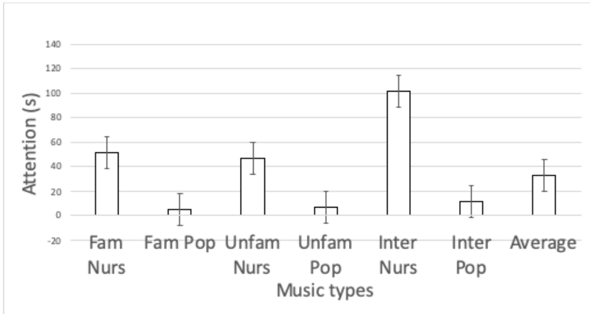


Figure 19b. 23-month-old group

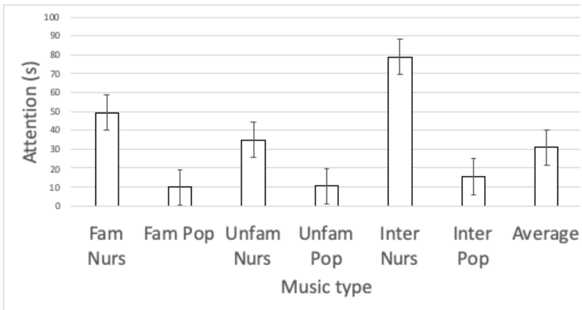


Figure 19c. All participants

Figures 19a-c. Average attentional bout duration elicited by music type. The latter comprise familiar nursery rhymes (FamNurs), familiar pop songs (FamPop), unfamiliar nursery rhymes (UnfamNurs), unfamiliar pop songs (UnfamPop), nursery rhymes through which parents had previously interacted with their children (InterNurs), and pop songs through which parents had previously interacted with their children (InterPop).

In terms of ceiling effects— whenever children listened to the music presented to them until the very end of the track —, results show that Daisy, Patricia and Gwendolyn are above the overall average with averages of 0.9, 0.6 and 1.6 respectively. Only Gwendolyn's excursion turned out to be statistically significant (t(12) = 2.39, p = .01).

a)	Hugh	Louisa	Daisy	Tommy	Patricia	Gwen	Average
Rejections	0.1	0.0	0.0	0.1	1.1*	1.6**	0.5
Ceiling effects	0.1	0.3	0.9	0.2	0.6	1.0*	0.5

b)	Hugh	Louisa	Daisy	Tommy	Patricia	Gwen	Average
Av att bt lgth	16.9	22.6	47.2	13.25	37.08	36.49	29.5
Group av		28.9			28.9		

Tables 29a and 29b. Table 29a. Average rejections, ceiling effects per participating child, and grand average (far right-hand side). Table 29b. Average attentional bout length paid to non-interactional music per participating child, and grand average (far right-hand side). * indicates statistical significance at alpha = .05, and ** indicates significance at alpha = .001

Unlike in the case of correlations between pragmatic development and maternal prosody, the relationship between maternal prosody and attention to recorded music did not evidence any patterns such as the ones described in Table 27. No particular register or pragmatic function evidenced significant correlations across participants, but unsorted IDSp as well as inquiring and pedagogic IDSp tended to concentrate them. On the other hand, all significant correlations between LUI raw scores and average attentional bout length are almost completely concentrated in the first age group (Table 30). Although no separate musical stimuli systematically generated statistically-significant correlations across participants, the first four participants evidenced so in at least one overall average (either including or excluding interactional songs, see two last columns in the table).

	Musical stimulus							
	FamNurs	FamPop	UnfamNurs	UnfamPop	InterNurs	InterPop	Av	Av-Inter
P1	-0.3	-0.6*	-0.8**	-0.5	-0.8**	-	-0.6*	-0.6
P2	-0.1	0.5	0.3	-0.3	-0.4	-0.3	0.4	0.4*
P3	-0.4	-0.5	-0.7**	-0.8**	0.1	-0.4	-0.5*	-0.7**
P4	-0.4	-0.5	-0.3	-0.6	0.5	-0.4	-0.3	-0.6*
P5	-0.2	-0.1	-1.0	-0.2	/	/	0.0	0.0
P6	-0.2	-0.1	0.3	0.4	/	/	0.4	0.1

* = p < .05, ** = < p < .01

Table 30. Pearson correlations coefficients between LUI raw scores and average attentional bouts, per participant, genre, familiarity, and whether they were associated with interaction or not (see cases 1 or 2).

One further angle from which to analyse attentional data is to observe the average length of attentional bouts directed at non-interactional music of all the children at the same time. Figure 20 thus displays a scatterplot containing the weekly average of each participant, in reference to its corresponding LUI raw score. The scatterplot evidences that no particular pattern can be observed between the two variables. Children at the same level of pragmatic development presented sometimes highly dissimilar average attentional bouts. For instance, around 20 points in the LUI scale, Louisa's average attentional bouts vary between 4 and 38 seconds in length, while Daisy's are over fifty. Daisy presented, by far, the highest average attentional bout durations in her age group (47.2s, see Table 29), and overall. In terms of age, even when she was slightly older than the other two participating children in her age group, her attentional bouts were longer from her very first week of assessment. Similarly, Patricia and Gwendolyn both present average attentional bouts much higher than Tommy's, even when he was older than the previous two participants at the start date of the study.

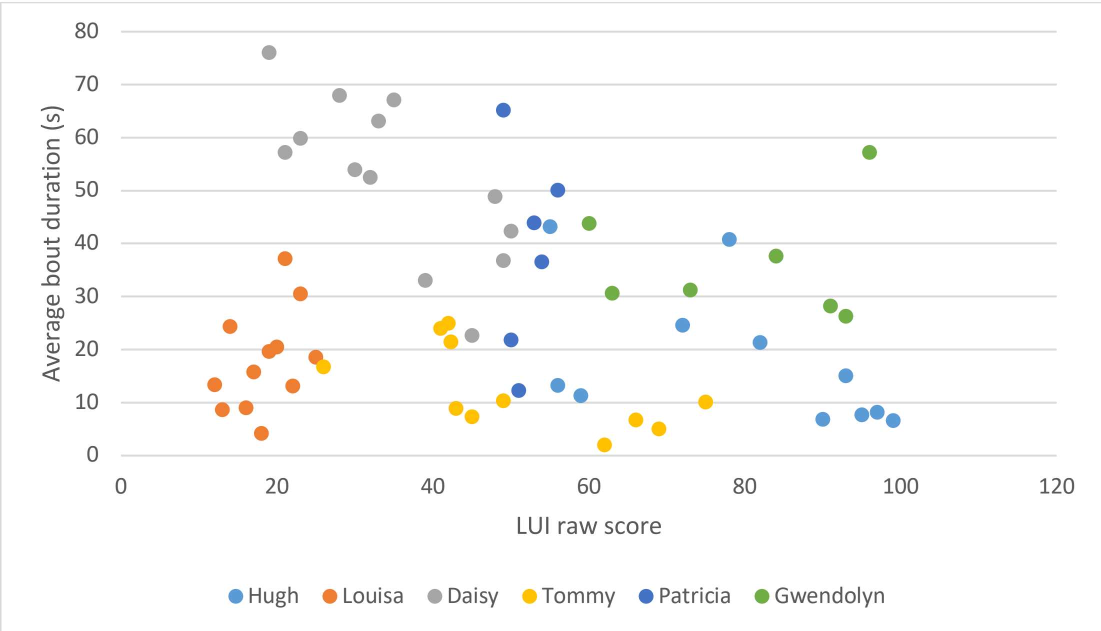


Figure 20. Weekly average length of attentional bouts directed at non-interactional music, per participant[28].

3.2.2.2. Attention to the parent

Further qualitative assessment was undertaken in an attempt to better understand what could lay beneath Daisy, Patricia and Gwendolyn's detachment from the other three participating children in terms of the length of their attentional bouts. One particular lead to

[28] Participating children's LUI raw scores did not always increased, and sometime repeated themselves. As a result, in very few cases, the same LUI raw score corresponded to two or more different LUIs (completed in different weeks).
In order to plot all participating children simultaneously and with a single, continuous scale of LUI scores in the x axis, whenever a given point in the scale corresponded to two to or more attentional data points of a participating child, these were replaced by their average.

follow was provided by the second study's discussion on cultural learning. In such a discussion, the possibility was contemplated that recorded music could constitute an object of attention, and that 18-14 months-old children pay attention to recorded music in an attempt to understand adults' interest in engaging with it. Unfortunately, cultural learning can only be empirically assessed by studying a child's gaze, a matter on which the second study provided no direct evidence whatsoever.

By focusing on children's gaze, qualitative assessment indeed suggested that children in the 15-month-old group that had not attended music playgroups, and the boy of the 23-month-old group consistently directed their gaze at their mothers as an initial response to the sound of recorded music. On the other hand, footage containing Daisy, Patricia and Gwendolyn's responses suggested that their gaze was more often than not directed at the music source itself.

Such a qualitative notion was quantitatively assessed by means of two measures. First, how long after acknowledging the presence of recorded music and its source did children first looked at their mothers. Second, what percentage of the song's duration did children look at their parents for. Regarding the first measure, Table 31 shows that indeed Hugh, Louisa and Tommy tended to look at their parents only a few seconds after the beginning of the musical stimulus (3.7s, 2.1s and 3.5s, respectively). On the other hand, Daisy, Patricia and Gwendolyn (12.7s, 10.1s and 29.3s) did so after over eighteen seconds, on average. Furthermore, there were cases when these three girls would not look at their mothers even once, but simply focus on the music. By aggregating the individual time measurements of Hugh, Louisa and Tommy, and contrasting them with the remaining girls' equivalent, a one-tailed Student's-t test confirms the dissimilar nature of both groups ($t(41) = -10.25$, $p < .0001$). As expected, a factorial ANOVA only returned a statistically-significant difference between Hugh, Louisa and Tommy as a group on the one hand and Daisy, Patricia and Gwendolyn on the other [F(1)= 4.02, p < 01].

Child	Average time (s) of first eye contact after song start	Group Average
P1	3.7	3.1
P2	2.1	
P4	3.5	
P3	12.7	13.5
P5	10.1	
P6	29.3	

Table 31. Middle column shows how long after acknowledging recorded music and its source did each participating child look at their mother. Top right column shows a further averaging of the previous scores.

Child	% of a song's duration looking at mother	Group Average (%)
P1	37	51.33
P2	51	
P4	66	
P3	13	14.2
P5	10	
P6	21	

Table 32. Middle column shows the average percentage of a song's duration during which each participating child looked at their mother. Top right column shows a further averaging of the previous

Regarding the second measure— what percentage of a song's duration did children look at their parents for—, Table 32 also suggests important differences between the groups mentioned in the previous paragraph. While Hugh, Louisa and Tommy spent, in average, around half of a song's duration monitoring their mother's behaviour, Daisy, Patricia and Gwendolyn did so in less than 15%. Once again, by aggregating the individual time measurements of children 1, 2 and 4, and contrasting them with the remaining girls' equivalent, a Student's-t test reveals statistically significant between groups (t(41) = -4.61, p < .001). A factorial analysis again confirmed a statistically-significant difference between Hugh, Louisa and Tommy as a group on the one hand and Daisy, Patricia and Gwendolyn on the other [F(5)= 8.26, p < 001].

4. Discussion

4.1. LUI and prosodic data.

4.1.1. Individual changes on weekly basis. Hypotheses 2.4.1 and 2.4.2

As supported by data, the LUI successfully assessed children's pragmatic development. In all six cases presented in this report, and even when assessments had been distanced one another on the relatively short interval of a fortnight, children's LUI scores increased quickly and steadily enough to be fitted into robust linear regression models.

As expected, LUI raw scores presented noticeable score increments between assessments in some cases (e.g. Hugh and Tommy). However, such increments were not consistently followed by changes in the parents' prosody. Out of the many prosodic parameters assessed through twenty weeks, only a number presented changes that seemed time-related to LUI scores. Furthermore, such changes were neither sustained in time nor repeated across participants (i.e. Hugh's mother's IDSp Speech Rate does evidence a sudden increase in week 9 (thus following the LUI raise on week 7), but this was not maintained throughout the following weeks (neither it kept increasing, nor it remained levelled), nor did it occur in any of the other cases. Taken together, these results suggest that peaks or dips throughout weekly assessments correspond to random variation (statistical error) stemming either from the complex nature of human development or— more likely —from an inevitable degree of inaccuracy in the subjective assessment of the mothers— reflected in the LUI questionnaires. Accordingly, it can be concluded that evidence gathered through this study does not support Hypothesis 2.4.1. In other words, distinct score increases between fortnightly LUI scores do not seem not to be followed by corresponding distinct changes in the parents' prosody, rendering both variables independent from each other in this particular regard.

This first conclusion can be verified by the fact that prosodic measure's sudden dips or increments were not systematically confirmed by significant correlations with raw LUI scores through time. Furthermore, these correlations' support for hypothesis 2.4.2 is only partial and inconsistent. First, significant correlations were, in general, not systematically found across groups or participants. Such a heterogeneity applies both to their number (11 in Hugh's case, 7 in Louisa's, 12 in Daisy's, 6 in Tommy's, 2 and 1 in Patricia and Gwendolyn's cases respectively) or prosodic item (e.g. ADSp, EnqIDSp, etc.). For instance, although f0- and contour-related measures were close candidates, no single prosodic item presented statistical significance systematically across participants. One notable exception to this overall lack of systematicity is unsorted IDSp. In all six studied cases, there was at least one prosodic measure in this register that reached statistical significance. This is an intriguing result and would prove relevant should it endure in a larger sample. A second manner in which correlations between LUI raw scores and prosodic measures support hypothesis 2.4.2 only partially and inconsistently is that they only sometimes represent a shift towards ADSp standards. For instance, contrary motion to hypothetical expectations, correlations often implied an increase in prosodic exaggeration as reflected in f0 mean or median. Thus, a second conclusion to be drawn is that data do not support that as LUI scores increase over time, the features of IDSp prosody systematically shift towards A-D prosodic standards on weekly basis (hypothesis 2.4.2).

4.1.2. Overall data and between-groups differences. Hypotheses 2.4.5 and 2.4.6

The second conclusion, and others that will come in this discussion, rest on the robustness of prosodic data. For instance, the reader might legitimately wonder whether the comparatively-lower number of correlations found in Patricia and Gwendolyn's cases primarily reflects their also comparatively-lower number of prosodic datapoints. At a more general level, it may be well suspected that the vocal data gathered in this study was not enough and thus skewed, in turn not reflecting the actual characteristics of the different registers (IDSp or ADSp) or pragmatic functions (Inquiring, pedagogic, etc). In such a case, the already presented assessments of hypotheses 2.4.1-2 would be compromised. However, as reported in section 3.2.1, figures 13a-c and 14a-c, both unsorted IDSp utterances, as well as sorted pragmatic functions were consistently more exaggerated in terms of f0 Mean, f0 Median, f0 Maximum, f0 Range and f0 Standard Deviation, than ADSp utterances. The available IDSp vocal data thus meets one of the essential criteria that has characterized IDSp since its inception: elevated pitch (Ferguson, 1977). As reported in Figure 17, IDSp vocal data also evidenced double the amount of f0 contour glissandi, rises and falls (as detected by Prosogram) than that found in ADSp. This is coherent with the established idea that f0 contours in IDSp are more exaggerated and smoother than in IDSp (Trehub, 2016). Interestingly, something similar can be concluded when observing nPVI data. As discussed in the Analysis section, it was expected that IDSp nuclei durations would overall be more even than those found in ADSp, where speech comparatively-less clearly uttered. Therefore, it was also expected that the nPVI of the former would be lower than that of the latter, and it is precisely what can be evidenced on Figure 16. Such a finding is also aligned with existing literature concerning nPVI differences between IDSp and ADSp (Lee et al., 2014) as well as between IDSi and ADSi (Hannon et al., 2016). By this token, as a whole IDSp data does behave as established in the literature, at least in terms of f0 and nPVI. The fact that speech rate and other time-related measures did not return expected patterns as f0-related and nPVI did, remains puzzling, and does not echo Ko's (2012) results. This could well be due to technical reasons, as previous studies have suggested that pitch-related descriptors can show better performance than time-related ones (Salselas & Herrera Boyer, 2011).

As previously reported, in both age groups ADSp remained largely distinct from the rest of the vocal data, with the lowest f0, the only exceptions being found in two of Tommy's mother's parameters (f0 Maximum and Range.) This outcome is relevant, considering that Patricia and Gwendolyn's mothers participated for a shorter period than the rest of the children, and that their mothers occasionally failed to send the weekly recordings (See Appendix I). These two cases' comparatively-smaller number of prosodic datapoints could have led to consistently diverging results. However, against such suspicions, it is actually Tommy's, and not Gwendolyn Patricia's mother whose prosodic parameters contradict the rest of the cases. The mentioned comparatively-smaller number of prosodic datapoints can also explain the fact f0-related measures in this group appeared slightly more clustered together than in the 15-month-old counterpart. Indeed, the average standard deviations of all of Patricia and Gwendolyn's f0-related measures were smaller than the ones corresponding to the other four other cases. Another difference seems to be that Patricia and Gwendolyn's f0 parameters are not as clearly detached from their mothers as the rest of the cases. These two girls' f0 Maxima is around 600 hertz, thus 100 lower than the other cases.

Interestingly, it was unsorted IDSp (i.e. all IDSp utterances collapsed, regardless of pragmatic function) that presented the highest f0 parameters. This is due to the fact that unsorted IDSp comprised all possible pragmatic functions beyond the four so far reported, including pragmatic functions that were ambiguous, or only seldom used and thus would not allow for weekly comparisons (e.g. negotiations, etc). It is evident from data that such minor/unclassified utterances contain the highest f0 parameters— and the lowest nPVI values. Amid the sorted pragmatic functions, Playful IDSP utterances were the most exaggerated, due to their characteristic higher levels of arousal. The sole pragmatic function the f0 parameters of which were sometimes lower than ADSp was Gratifying IDSp. The explanation for the latter can be found by listening to such utterances: in order to gratify their children, mothers almost always resort to culturally-defined phrases such as ‘good girl!’ or ‘good boy!’. These phrases naturally comprise f0 contours that are no less culturally-defined, and present descendant glissandi. As a result, an exaggerated gratifying utterance would go lower than normal in terms of f0 contour, just as an exaggerated question contour would go higher. Such notion is confirmed by Figure 20, which shows that IDSp Gratifying and Inquiring utterances present the largest amounts of glissandi, when compared the rest of the assessed pragmatic functions and registers.

Present data goes beyond ratifying existing literature and provides novel— if preliminary —sketches of prosodic profiles of IDSp used in more specific, distinct pragmatic functions (Figure 15). As previously mentioned, the study of IDSp has, to a large extent, suffered of a somewhat ‘monolithical’ approach to its object of study (Farran et al., 2016). A notable example of such a lack of nuance is the fact that most existing experimental work does not control for affective differences between IDSp and ADSp stimuli (Singh et al., 2002). The latter is important given that systematically contrasting positively-emotive IDSp stimuli with more ‘matter-of-fact’ ADSp stimuli would generate a confound between affect and register. Nuance to IDSp prosody has been provided by studies that have evidenced how f0 contours change according to different contexts (e.g. ‘approving’ or ‘disapproving’) (Papoušek et al., 1990) or intentions (e.g. expressing approval, soothing, etc.) (Fernald, 1989, 1991). The present work thus contributes to such nuances by focusing not in contours, but in a broader series of measures, both f0- and time-related. As reported, the different IDSp pragmatic functions’ relative rank of exaggeration depends on the measure observed and are (obviously) not quite as distinctly detached from each as registers (IDSp (green line) v/s ADSp (sky-blue)) or speakers (the child’s vocalizations (blue line) vs all the rest). The application of this analysis to larger datasets could confirm whether these prosodic profiles are indeed distinct or will collapse.

Having verified and established the validity of both groups' vocal data as a whole in terms of speech registers, it is now possible to discuss between-group differences. One such difference is that two of the three cases in the 23-month-old group (Patricia and Gwendolyn) presented lower f0 parameters than the rest of the children (Figures 18b and 18c vs 17a-c and 18a). Whereas the latter's f0 maxima reached or surpassed 700hz, Patricia and Gwendolyn‘s did not surpass 650hz. Remarkably, Patricia's f0 parameters, although comparatively-high, are not distinct from her mother's, as if comprising a more adult-like prosody. These findings only partially support hypothesis 2.4.6, insofar as that inter-group differences in terms of the children's prosody can be found.

In terms of maternal prosody, a few relevant results deserve to be discussed. As discussed in the previous section, the number if significant correlations between LUI raw scores and maternal prosody were not systematic found across participants (11, 7, 12, 6, 2 and 1, in the cases of Hugh, Louisa, Daisy, Tommy, Patricia and Gwendolyn, respectively). Nevertheless, when looking at the situation from a wider angle rather than the individual level, a trend can be noticed: the number of significant correlations decreases with age, making both groups differ in their averages. These results provide support for the idea changes in maternal prosody are not even throughout children's linguistic development. In line with the VAW hypothesis, they constitute an example of how the relationship between changes in children's linguistic development and their mother's prosody are more tightly bound around 18 months of age (first group) than around the 24 (second group). Findings discussed in this paragraph partially concur with Ko (2012) and Farran's (2016) in depicting of the end of the second year of life as a period comprising important changes in terms of maternal prosody, but place the occurrence of changes rather at the earlier side of this time window— around 18 month of age. In any case, the small sample size of this comparison does not allow for any kind of generalization, and these results should be taken as preliminary until being tested against a larger sample. For these reasons, inter-group data does not strongly support hypothesis 2.4.5.

The final result to be discussed in this section concerns pedagogic IDSp. As discussed in chapter 3 (section 2.2.2.1), the theory of natural pedagogy (ToNP) illustrates how the transmission of knowledge from an adult to a child is made explicit by means of non-verbal, ostensive cues such the use of IDSp. As part of its capacity to draw the infant's attention, IDSp is capable of triggering a learning disposition in the child. In this respect, the present study segregated IDSp sentences according to pragmatic functions— including pedagogic IDSp. Although without originally bearing in mind the ToNP, pedagogic IDSp utterances consist precisely of instances when IDSp was being used as an ostensive cue for signalling the transmission of knowledge. This is not entirely a coincidence, as pragmatic functions were selected according to the researcher's impression of the most pervasive pragmatic uses of maternal speech in the corpus. Results in Table 27 evidenced that the older the child was, the more likely it was that their mother would use IDSp as an ostensive cue, with children in the second group receiving around 75% more pedagogic IDSp sentences. Although once again only preliminary, this result is also relevant as it demonstrates a consistent shift in the use of IDSp as children grow older. They also provide some of the first longitudinal evidence of the development of ToNP as describe by Csibra and Gergely (2009).

4.2 Attentional data

4.2.1. Attention to music. Hypotheses 2.4.3, 2.4.4 and 2.4.7

The rejection of hypothesis 2.4.1 summed to the fact that no patterns were found on weekly basis in terms of attention to music remove grounds for the support of hypothesis 2.4.4 (distinct changes in the mother's IDSp prosody will be followed by distinct increases in terms of the length of the attentional bouts that children pay to recorded music).

Data's alignment with hypothesis 2.4.3 (negative correlations between the mothers' IDSp features and the length of attentional bouts that children pay to recorded music; as IDSp shifts towards A-D prosodic standards, bouts will become increasingly longer) was mixed. As mentioned in section 4.1.2, data in tables 6, 10 and equivalent, containing correlations between maternal prosody and attention to recorded music did not evidence any patterns, and no particular register or pragmatic function presented significant correlations across participants. In terms of the temporal unfolding of children's attention to recorded music, Louisa's attentional bouts tended to increase, as predicted by hypothesis 2.4.3. This finding, summed to certain prosodic measures that also behaved as expected, provides some support for such a hypothesis. However, in the rest of the cases, the average length of attentional bouts tended either to decrease (Hugh, Daisy, and Tommy) or to remain constant (Gwendolyn and Patricia), thus not behaving as expected. When observing the average length of attentional bouts directed at non-interactional music of all the children at the same time (Figure 20), no particular pattern could be observed either between the two variables. Taken together, these findings do not support hypothesis 2.4.3.

Regarding hypothesis 2.4.7 (inter-group differences in terms of the length of the attentional bouts that children pay to recorded music), data provides a number of interesting findings. As presented in Table 29b, the average length of attentional bouts elicited by non-interactional music is virtually the same in both groups. Such a homogeneity is explained on the hand by Daisy's high and Tommy's low averages. Another relevant finding is the exponential increase in the number of rejections and ceiling effects shown by Patricia and—especially —Gwendolyn. Daisy (a member of the 15-month-old group) presented a high average of ceiling effects, but a post-hoc interview with her mother revealed that it was most probably due to her attendance to Children Music Groups, which fostered her habit of attending to nursery rhymes. Patricia and Gwendolyn on the other hand, although not belonging to any Children Music Groups, proved interested in paying attention to the whole of a nursery rhyme from the first sessions onwards. The latter creates an evident contrast with the rest of the children, the few ceiling effects of which started occurring after several recorded sessions.

In the same fashion, right from the start of the study Patricia and Gwendolyn seemed to have already developed clear preferences in terms of music genre, as they would manifest strong rejections to some of the pop music presented to them. As an extreme illustration, Gwendolyn rejected the pop song chosen by her mother to interact through. Her rejection remained unaltered over the next two weeks, so the researcher and her mother agreed to stop trying to use that particular song. Such findings strongly support hypothesis 2.4.7, as a number of inter-group differences in terms of the children's attention to music are found.

Findings also suggest an incidence of sex and/or gender. It is only girls (Daisy, Patricia and Gwendolyn) that evidence average attentional bout durations double those of the rest. It would be interesting to see if such differences remain when assessing a larger sample, as little girls have been reported to develop faster than boys in a number of key scholastic attainment areas (Bourke & Adams, 2012) including linguistics (Faraj, 1988), and as ratified by general higher performance of girls in the LUI percentile tables (O'Neill, 2007).

In addition to the quantitative and statistical assessments performed and summarized, upon observing the ORViRT sessions it was the researchers' clear qualitative impression that Patricia and Gwendolyn — already by the time the study's first sessions took place — engaged with music in a manner different from that of the four other participants. In a way, it seemed like they had developed a certain personal taste for music. They almost seemed to be able to quickly discriminate whether a song pleased or displeased them and would accordingly easily remain focused on them for as long as the song lasted, or quickly and explicitly reject them. On the other hand, although Daisy engaged in some of the longest attentional bouts, she was not as active or selective and paid attention to music more indiscriminately (see Table 28). Interestingly, Patricia and Gwendolyn presented a number of common, quantifiable behaviours. Both of them presented at the same time a) f0 parameters different from the rest, b) longer average attention and c) more ceiling effects and rejections than the rest of the cases. Thus, as a whole, these results do provide support for hypothesis 2.4.7.

Results can be understood beyond averages. Children's attention to the different types of music did describe clear overall configurations. As summarized in Figures 19a-c, genre (nursery rhymes or pop) appears to be the most relevant aspect when it comes to drawing children's attention. Children evidenced attentional bout durations significantly above average as a response to familiar nursery rhymes, and even more so to nursery rhymes associated to interaction with their mothers. On the contrary, familiarity and interaction were not enough to elicit children's attention to pop music to levels even close to the average. These differences in terms of genre and association with interaction, can be partially understood in terms of the enculturating relevance they each may carry. Indeed, parents often use the conventions embedded in nursery rhymes or children's songs as means to nourish their children's understandings of cultural conventions; in such contexts, shared musical activities may function as 'signs' or 'cultural tools' (Williams et al., 2015). Thus, on the one hand, parents bestow on music' lyrics and actions an interest that goes beyond aesthetics or affectivity. On the other hand, children's mastering of these musical elements is accordingly added a pedagogic layer of reward, and further motivation.

These notions are thus reflected in the descriptive statistics, as well as in the qualitative appraisal of the videos. An illustrative example is Louisa's attention to 'If you're happy and you know it' (the song her mother chose as InterNurs), which was indeed the highest she paid to anything else. In the videos, her reactions do not only consist of expressions of positive affect such as smiles, but of appropriate actions that were also enthusiastically celebrated by her mother. This is particularly remarkable considering that (as specified in a post-hoc interview) such actions had only been introduced to her for the first time literally a couple of weeks before. Another illustrative example is that of Tommy. As previously mentioned, his reactions to recorded music often consisted of short, rather stereotypical dancing or swaying. His reactions to 'The wheels on the bus' (the song his mother chose as InterNurs), on the other hand, were utterly contrasting in terms of both length and intensity. As previously described, Tommy would not perform actions carefully and rather meticulously, but would also protest whenever his father tried to join the performance as if actions were his 'exclusive' business. Such a behaviour conveyed a sense of identification with the setting and activity, as if he already understood that 'nursery rhymes are for children'. In other words, existing theory and current evidence seem to converge in suggesting that children like and devote attention to nursery rhymes and children's songs because there is something for them in such genre, something they can easily relate to and engage in. This was especially true for children in the first age group but could also be evidenced in the second.

The mentioned lack of interest for pop music seemed only to increase with age, as evidenced when comparing figures 19a and 19b. In all cases, pop music (with or without associated interaction) was less attended to than nursery rhymes (with or without associated interaction). The contrast in terms of children's interest in these two genres seems to echo previous literature indicating that parents sing far more children's songs and lullabies than pop music (Custodero, 2006; Custodero & Johnson-Green, 2003). The appeal and cultural relevance of genres that have children as a specific target are once again highlighted by the fact that in the present study participating children consistently paid more attention to unfamiliar children's songs or nursery rhymes (UnfamNurs) than to pop music they were familiar with (FamPop) or had interacted through (InterPop). This is remarkable given that children could not have related them to their identity or past experience (Barrett, 2010). Furthermore, the mentioned songs were rendered entirely in foreign languages (French and Spanish), ruling out possible associations in terms of actions or lyrics. One possible explanation for their appeal lies in the non-verbal, purely-musical elements of nursery rhymes and children's songs, such as melodies and the timbre of musical instruments. It was the researcher's qualitative impression while examining the videos, that the mere tempo, rhythm and timbre of some instruments were enough to capitalize children's attention, sometimes after the very first bar of music. These findings suggest an universally-effective acoustic appeal of children's songs, such as that already studied in the case of lullabies (Trehub et al., 1993a, 1993b).

4.2.2. Attention to the parent and cultural learning

Another interesting finding concerns— as suspected following the second study's results —cultural learning (Tomasello et al. 1993) and epistemic trust (Fonagy et al., 2015). In the last discussion, the Random Forest analysis positioned interaction with a significant other as the single most predicting factor for sustained attention to recorded music. It was thus hypothesized that children's gaze would play a major role in confirming the role of cultural learning in children's sustained attention to recorded music. Findings presented in section 3.2.2 seem to confirm such hypotheses. Indeed Hugh, Louisa and Tommy looked at their parents only a few seconds after the beginning of the musical stimulus. Because children seemed to take at least one second to recognize the aural stimulus as music, it seems safe to deduce that looking at their mothers for orientation was the very first action these children took. Participating toddlers would often wait for their mother's smile before smiling. They also seemed to only shyly sketch dancing movements and proceed to properly dance only once receiving encouragement or example from their mothers. In the case of Tommy (older than Hugh and Louisa), observing his parent seem to focus particularly on learning how to sing along and do the exact actions of some of the nursery rhymes. Thus, findings provide interesting— if preliminary — support for the idea that it was cultural learning in general, and imitative learning in the domain of object-directed action in particular (in this case, culturally-appropriate responses to recorded music such as smiling, dancing and singing along) that was taking place. At the same time, these children's immediate seeking for their parents in the context of facing recorded music stands as an instance of epistemic trust. As discussed, a cultural product such as recorded music involves a degree of opacity in the context of which the caregiver's status as a privileged, trusted source of knowledge becomes evident. In Hugh, Louisa and Tommy's case, it seems like music remains still— at least partially —opaque, and how to 'use it' is not evident for these children (Is this music for dancing? Is this music that has associated actions? Is it music I should smile to? etc.) Thus, their rather immediate need for eye contact seems to illustrate in turn their resorting to their mothers as sources of guidance and confirmation regarding 'what to do' when faced with recorded music.

At the same time, results also showed that Patricia and Gwendolyn (both girls in the 23-month-old group that had not attended music playgroups), as well as Daisy (the girl in the 15-month-old group that had attended music playgroups) sought their mothers' gaze considerably less and later. In their case, attention was paid to music directly, and behaviours such as smiling, and dancing were performed comparatively more spontaneously and autonomously. Such an autonomous appraisal of music seems to correspond to cultural learning's second step, instructed learning, through which children internalize the instructions of their role models and use them subsequently to self-regulate their own attentional, mnemonic, or other cognitive functions, to later re-enact the adult's input overtly in regulating their own behaviour when faced with the same or a similar situation.

The mentioned behaviours and impressions can equally be understood in terms of mentalization. Hugh, Louisa and Tommy's need for eye contact with their attachment figures suggest a lack of mentalisation of their own when approaching music, and a resulting need for their caregiver's mentalisation. Indeed, it seemed as if these children needed their parents to reflect on their experience, to have 'their mind in mind' (Fonagy & Campbell, 2015), which in this case involved the mother's acknowledgment of the children's own acknowledgment of music as such and only after that a confirmation of their reactions (smiles, actions, etc.) as appropriate responses. On the other hand, Daisy, Patricia and Gwendolyn's long bouts of silent, autonomous attention to musical stimuli suggested a much more developed mentalization capacity. The latter allowed them to experience music on their own as if they had successfully internalized what the previous what the other three children were still internalizing (music-related appropriate reactions)— as if they already had their caregiver's mind in mind.

The case of Daisy, the girl in the 15-month-old group that had attended music playgroup, illustrates the considerable facilitating impact in musical enculturation that such activities can provide. Even though her pragmatics were not the most advanced to be found in her group, her comparatively-high average attentional bouts, ceiling effects and visual behaviour converge in depicting a degree of autonomy in her relationship to recorded music that resembles more that of children up to 6 months older.

4.3.Integrated discussion

The sum of the assessed behaviours and qualitative impressions strongly suggest that important changes concerning participating children's motivation and capacity for paying attention to recorded music do occur during the developmental window scrutinized in this study and provide insight their relative importance. However, available data contradicts most of the hypotheses in terms of the impact of maternal use of IDSp on children's attention to recorded music, the nature of the changes that IDSp, and the associated temporal scales. Overall, data do not support the idea that such changes in children's engagement with recorded music are directly driven by changes in the mother's prosody, or the children's linguistic and pragmatic development on weekly basis. Instead, as in the case of the previous two chapters, the current study underlines the importance of social, emotional and cultural factors such as mentalization, cultural learning, natural pedagogy (ToNP). Crucially, the design of the study and its manipulation of social interaction seems to have led to the manifestation of the latter's paramount importance. Some changes involving pragmatic competence and maternal IDSp do seem to play a role, but the latter is only visible when contemplating longer time periods— comparing groups separated months apart, instead of following weekly changes. In the next paragraphs, a review of the study and its methodology will serve as a guiding line for conclusions.

As explained in this chapter's introduction, a setting was needed where children could spontaneously— but would not be externally prompted to —engage with recorded music; where recorded music would be casually presented and without further stimuli, and it could be directly assessed whether the child would pay sustained attention to it, and whether they would spontaneously wish to supplement listening with forms of musical play. Accordingly, a one-way remote video recording technique (ORViRT) was designed in order to gain access to such a natural setting while introducing as few unusual elements as possible. Such a goal was only partially reached, and a number of the present results can be attributed to different aspects of the applied methodology.

Parents as a whole did not deem the setting they were asked to generate once a week as utterly artificial or unnatural in an undesirable manner. Unlike the rest of the parents, Louisa's mother was the only one who chose for the sessions not to take place in the living room. For most parents, thus, the situation of music being played while they were both sharing the living room space but doing separate activities was a familiar one. Even if Daisy's mother qualified the setting as 'unnatural', Daisy did not seem distressed at all in the videos, even during the ones corresponding to the first weeks. Nevertheless, most parents were nervous— as was the researcher —at least during the first few weeks and as a consequence, children were too. A symptom of such anxiety-inducing effect of the first sessions is the fact that a number of participating children threw temper tantrums during one or more of the first couple of sessions while doing so less consistently— or not at all —in later ones.

One aspect of the designed methodology that had vast implications was the chosen source of music: smartphones. Most temper tantrums were consequences of quarrels between parents and infants over such devices. Indeed, almost without exception participating children searched for the source of the music being played, and often felt compelled to manipulate it or take it away. Parents often frustrated such attempts, partially in order to keep the music playing and partially to be able to remain in contact with the researcher via WhatsApp. Quarrels may have eventually discouraged not only children's (e.g. Hugh) trying to access the devices, but also their attention to the music being playing through them. Quarrels over who can or cannot access a device constitute perfect examples of 'terrible twos' situations, where toddlers are exploring and testing the limits their control over themselves and others (Green, 2000). By this token, quarrels and temper tantrums constitute an integral part of children this age and their parents and, even if they momentarily hindered the methodology's goals— observe children's reactions to recorded music, should not be regarded as excursions from a natural setting.

Another major consequence of using smartphones as a source of music was that they made it sometimes impossible to determine whether it was music or the screen children were paying attention to, thus generating ambiguity and eventual data loss. Even if music tracks had no moving images associated, screens seem to have often been compelling in themselves. Out of these experiences, it seems like much would be gained in next versions of the present methodology by incorporating a third device: a wireless speaker. Settings counting with such an addition would allow for a better-suited division of functions: a laptop for recording, a smartphone for providing links and communicating with the researcher, and a speaker to play the music. The latter would entail no screen and could even be placed so as to be unreachable by the child.

Perhaps the element of the implemented methodology that carried the most significant consequences was the instruction given to the parents of refraining from influencing their child's reaction to music, its significance lying in the degree to which it prevented interactions from taking place that would have normally occurred. Such an instruction, let us remember, obeyed to the analytical approach of the present study. Because children are highly motivated to obtaining their caregivers' attention, as soon as a parent would use the music being played as a means for interaction (e.g. prompted the child to sing, dance, etc.) children would most probably engage in musical behaviour. However, it could not be said that it was music itself that drew the child's attention, but rather the adult's invitation.

The neutrality instruction fulfilled its purpose, as children's spontaneous reactions to the vast majority of the musical stimuli could be observed. Nevertheless, it also added artificiality to the situation, as parents normally often react spontaneously to music themselves, directly or indirectly inviting children to join in. The artificial character of such neutrality seems to have had a direct impact on the average length of children's attentional bouts. As detailed in this chapter's introduction (section 2.3.4.2), participating dyads interacted to the sound of one particular nursery rhyme (InterNurs) during the first seven weeks of the study and without being observed (not during the sessions). From the eighth week onwards, each participant's chosen nursery rhyme or children's song was played once during each session. As evidenced throughout the individual cases and on the exception of Tommy, all children presented a degree of increment in their average bout lengths around the addition to the sessions of the piece to the sound of which they had been interacting off-camera. Such increments make sense: songs selected as InterNurs were freshly associated with significant interaction, and children re-enacted on-camera what they had systematically enjoyed for over a month.

What followed week eight was no less interesting. InterNurs songs played during the eighth and following sessions differed radically with the previous, off-camera instances in that parents were no longer allowed to actively invite children into musical interaction. At the same time, Hugh, Daisy and Tommy presented a degree of decrease in their average bout lengths following that week. The length of Hugh's average attentional bout durations decreases rather steadily over the following month and a half, without ever returning to previous levels. Tommy presents a more sporadic dip about three weeks later than the introduction of InterNurs. Daisy's average attentional bout also starts decreasing after a few weeks. It thus seems like there was some sort of inertia, as children remained interested in the songs for a certain number of weeks. However, it also seems quite probable that participating children eventually perceived the contrast between before and after week eight. Because parents were no longer playing an active role around the selected songs, the appeal of such songs eventually waned. Qualitative observation of these three participants' footage gave the researcher the impression that children eventually realized that there was nothing as interesting for them anymore in these songs, nor something new or challenging, and thus opted by other, surrounding sources of interest (e.g. toys or other objects). A different case is that of Louisa, who evidenced a comparatively-lower linguistic, pragmatic and motoric skills. For her, it seemed like the exploration of social interaction through music was something less familiar and therefore fuller of challenge and discovery. As a result, she did not present signs of demotivation and continued to seek actively seek for interaction, and her average bout length tended to increase over the weeks.

Thus, the systematic combination of the neutrality instruction, the off-camera interaction, and changes introduced on the eighth week resulted in a veritable unintended manipulation of the interactional factor. Accordingly, results depict and highlight the direct impact of the social dimension of music, and to what extent it lies beneath children's motivation to pay attention to recorded songs. Such a direct impact becomes all that more evident when systematically manipulated, as was illustrated in the previous paragraphs. It can be hypothesized that a significant portion of the hereby reported attentional bouts would have been longer, would the parents have been free to interact. Such hypothesis could be empirically tested, and its results would provide a considerably more accurate image of the impact of interaction on attention to music. The importance of interaction is further illustrated by the fact that Daisy, the only participating child that had been consistently exposed to Children's music groups (and the meaningful interaction in them), presented the highest overall attentional bouts.

The available data also discards the idea that such a change might be directly driven by changes in the mother's prosody traceable on weekly basis. More importantly, rather than indicating that mothers' use progressively less IDSp in the studied developmental windows, or that the different aspects of IDSp (speech rate, f0 measures, etc.) change systematically, data suggests that the relative incidence and importance of the different pragmatic function of IDSp change over time. In particular, results evidenced a substantial difference between groups in terms of the mother's use of IDSp as an ostensive cue (pedagogic IDSp). This result is in line with previous discussion about mentalisation: as children in the second group (mainly Patricia and Gwendolyn) had further mastered verbal communication and its communicational advantages, their development of mentalization was accordingly higher. A third, congruent finding concerns correlations between LUI raw scores and average bout lengths. While participants 1 to 4 presented at least one significant correlation (either corresponding to the overall average or the non-interactional average), Patricia and Gwendolyn did not present any. This can be interpreted as their levels of mentalization being more established and developed to a point that made their engagement with music more stable.

As a whole, data suggests that changes in children's motivation and capacity for listening to music are mainly driven by social interaction. It also suggests that such changes heavily depend on sex and/or gender, and that they are related to the qualities and development of their own speech, rather than that of their mothers.

GENERAL DISCUSSION

In this final chapter, I will proceed as follows. I will start by presenting one last time this dissertation's two main empirical findings— concerning VAW and interaction, respectively —and discuss how they largely contradict the thesis presented in chapter 5. The first main finding is that the studies did not support the existence of VAW. As an explanation, I will account for what I believe I failed to observe when formulating the thesis: uses of IDSp beyond the affective sphere, such as the one described in the ToNP. The second finding— the role of socio-cultural interactions —is something I chose not to focus on but kept emerging from the data.

I will then adopt a more panoramic view of the findings and describe what I believe is the reason why I did not foresee such mismatches: a so far unacknowledged and unwarranted emphasis of 'centripetal' over 'centrifugal' attachment dynamics. This will be followed by an updated review of factors mediating toddler's attention to recorded music, and an explanation of why I believe that the analytical approach that characterized important parts of this dissertation was worth adopting. Further criticisms to this work will be presented, and further future directions for research outlined. Finally, conclusions will be drawn.

1. Two main findings

As explained in chapter 1, a continuity between the IDReg and subsequent engagement with Western recorded music has to date been largely taken for granted yet not explained. Malloch and Trevarthen's, as well as Trehub's theoretical approaches to the IDReg and later musical engagement draw a more or less tacit causal relation between these phenomena: that the unfolding of the former stands as a prerequisite for the latter to take place. Yet, as discussed (chapter 5, section 2), while early interaction and interaction within the IDReg take place during the first years of the child's life with a peak around 6 months of age, the most common form of Western music engagement (listening to a music recording) requires further enculturation and is not consistently observable until the end of childhood/early adolescence. By this token, a mechanistic, observable chain of events linking the two phenomena is tacitly assumed but remains virtually unproblematized and undescribed. Accordingly, in this dissertation I provided a first attempt at describing a mechanistic chain of events. In the main thesis, the theory of early musicality and later listening to recorded Western music were connected by using as a link the variation through time of the caregivers' use of IDReg: Vocal affective Weaning.

1.1. VAW

Data collected through this dissertation's third— and most important —study suggest that important changes concerning participants' motivation and capacity for paying attention to recorded music do occur during the developmental window scrutinized in this study, providing insight into the relative importance of parental use of IDSp and other concomitant relevant factors highlighted by the literature, such as synchronous interaction. Notwithstanding, the sum of the assessed behaviours and qualitative impressions (not restricted to the third study) contradicts to a large extent the thinking put forward in chapter 5.

Regarding the relationship between parental use of IDSp and changes in child sustained attention to recorded music, already the first study returned no testimonies or impressions in its direction. Instead, participants reported an increase in the expression of negative emotions such as anger as a direct consequence of acknowledging language acquisition, as well as evidencing its use for pedagogical purposes. In the second study's design, a theoretical commitment and a leap of faith were made in relying strongly on the quantal changes reported by the literature in terms of caregivers' use of the IDReg between 18 and 24 months of age. Indeed, following the main thesis, it was assumed that quantal improvements in the children's' linguistic development would imply a reduction in parental use of the IDReg. However descriptive statistics and the Random Forest analysis revealed no connection whatsoever between any of the scales contained in the MacArthur-Bates Communicative Development Inventories (CDI) and sustained attention to recorded music. A third, longitudinal study was therefore devised, that would directly and precisely assess children's sustained attention to music, pragmatic competence, and parental prosody. Although the Language Use Inventory (LUI) scores did present a number of rather scattered, unsystematic quantal increments between assessments, such increments were not consistently followed by changes in the parents' prosody. Something similar was evidenced at the level of parental prosody.

In the third study, parental prosodic measures could be divided into f0-related— f0 Mean, Median, Standard deviation, Maximum and Range —and time-related measures— speech rate, syllable nuclei duration nPVI, average rate in utterances per minute, and ratio between the total number of IDSp and ADSp utterances (IDSp/ADSp). In terms of f0-related measures, participants in the 23-month-old group presented lower f0 parameters than the rest of the children. Such results were expected, following previous research such as, for instance Ko (2012). These between-group differences were also expected to be somehow reflected at the weekly-basis level; however, it was not the case. As reported (chapter 8, section 4.1.1), only a portion of the weekly prosodic measures presented significant correlations in the expected valences (positive or negative), such changes being neither repeated across participants nor sustained in time, thus evidencing no link to 'quantal' changes in the children's pragmatic development (LUI scores).

In terms of time-related measures, an equivalent outcome was found. This was particularly unexpected (and to some extent, disappointing) in the case of time-related measures such as average rate in utterances per minute and IDSp/ADSp utterance ratio, which were measured strictly following Farran's (2016) guidelines. Given that when aggregated speech data overall behaved as expected (section 3.2.1 in the previous chapter), technical issues at measurement can hardly be held responsible. A more plausible reason could be that the targeted prosodic changes develop too slowly to be apprehended on weekly basis, as both Farran (2016) and Ko's (2012) studies spanned longer time windows than the one hereby assessed (see chapter 5, section 3.3.2). Another interesting point to consider is the situational dimension. In Farran's study, participating parents were instructed to interact with their infants as they normally did at home, without any mention of what the interactions actually consisted in. This omission is relevant since interactions could consist of routines or forms of interaction other than the ones included in the previous chapter.

F0- and time-related measures were not the only ones discussed in chapter 4 (section 3.2), as I argued that literature bestowed significant importance to IDSp's 'loving tone'. In this regard and although it was decided not direct resources to its systematic assessment, it seems appropriate to briefly discuss the matter in light of collected evidence. As discussed, f0- and time-related measures remain partial measurements of IDSp, not directly informing about its affective quality. In other words, it was unclear whether the former would be accompanied by the latter, at least in the selected developmental window. In this regard, parents' impressions collected in the first study rather converged in suggesting that their use of the 'loving tone' endured their toddlers developmental changes and the need for normativity. In other words, while the use of IDSp seems to naturally privilege the uninhibited expression of 'positive' emotions, it also seemed to allow the embed expression of frustration or authority into a 'safe transactional tone' (chapter 4, sections 3.2 and 4). In the same vein and for what it is worth, my own impression after listening to the sum of the weekly recordings reported the third study is that parental speech was no less affectionate in the 23-m-o group (compared to the 15-mo counterpart), nor did it seem to wane throughout the weeks. This would be consistent with previous findings concerning IDSi (another expression of the IDReg) showing that as children grow older parents change some acoustic features (for instance, generally lowering f0) while preserving the 'loving tone' (Bergeson & Trehub's (1999). Nevertheless, these impressions (participating parents' and my own) remain unsystematic and should be systematically assessed before drawing any reliable conclusions.

Thus, speech data results do not support the occurrence of VAW, at least as conceived in chapter 5. In turn, because distinct changes in the parents' IDSp prosody did not seem to take place, assessing whether they were followed by distinct changes in the children's attention to music was not quite possible.

Fortunately, data does not invalidate the VAW hypothesis without also providing insight as to why this seems to be the case. The most important reason for this mismatch between the main hypothesis and data has to do with the initial understanding of IDSp that the formulation rested on. Such a formulation put excessive emphasis on the affective aspect of parental speech (i.e. VAW) while disregarding others. Indeed, perhaps the most important flaw in the original formulation was the failure to take into account the multiplicity of functions that IDSp simultaneously serves, notably its use as an ostensive cue as described by the ToNP (Csibra & Gergely, 2009). As discussed (chapter 4, section 2), the ToNP illustrates the way in which IDSp can trigger in the child a learning disposition towards new, relevant information. Such a use of IDSp was widely— and spontaneously (it was not contemplated amid the focus groups' set of questions) —acknowledged by parents in the first study. Furthermore, the 8th chapter's speech analyses strongly suggested children in the 23-mo group received around 75% more IDSp in its use an ostensive cue than their younger counterparts. Albeit preliminary, this result seems to be the first of its kind in providing precise longitudinal evidence of the temporal unfolding of IDSp as an ostensive cue. Such an increase through time can be interpreted as a result of language acquisition: the more children understand words, the bigger the volume of verbal information that can be transmitted to them. So far, literature has focused on the use of IDSp as an ostensive cue at discrete ages in infancy; 6-month of age (Hernik & Broesch, 2019; Senju & Csibra, 2008), 8 (Wu et al., 2014). In contrast, present results directly compare its use across age groups, while at the same time covering toddlers instead of infants. Thus, results from the third study suggest that as children develop and acquire language, ostensive cues become more important and therefore more prevalent. This cognition-related aspects in fact seems to become progressively important while the affective dimension may— or may not —recede. Further research should scrutinize in further detail the relationship between language acquisition and parental use of IDSp as an ostensive cue.

1.2. Interactional aspects of music listening. Cultural learning and epistemic trust

The ToNP is, in turn, intrinsically connected to more than one aspect of attachment, such as mentalisation and children's use of attachment figures as secure bases for exploration, which can provide motivation for engaging in less familiar activities or with less familiar objects. As discussed at several points throughout this dissertation (e.g. chapter 1, section 4; chapter 3, section 1.3), recorded music is a cultural product that presents challenges to infants and toddlers in terms of how to approach it or what to expect from it. Indeed, recorded music can be more *opaque* to a child than IDSi or other forms of typical musical interactions described by in chapter 5, section 3.5 (e.g. Custodero, 2006; Marsh et al., 2016; Williams et al., 2015). It is no surprise then that studies 2 and 3 returned an important role of socio-cultural interactions, in children's motivation and capacity for attending to recorded music: epistemic trust, mentalization, and cultural learning.

As discussed in the last chapter (section 4.2.2), children in the 15-mo group often sought for their parents in the context of facing recorded music, which could be interpreted as an instance of epistemic trust. A cultural product such as recorded music involves a degree of opacity (e.g. Is this music for dancing? Is this music that has associated actions?) in the context of which the caregiver's status as a privileged, trusted source of knowledge becomes evident. Video analysis suggested that the younger children's rather immediate need for eye contact illustrated their resorting to their mothers as sources of guidance and confirmation regarding 'what to do' when faced with a partially opaque cultural product such as recorded music. Originally conceived in the context of psychotherapy (Fonagy et al., 2015), the construct has proved a useful tool for understanding certain aspects of attachment (Luyten et al., 2017) and pedagogy (Fonagy & Campbell, 2015). In the same vein, the present work provides a first illustration of the role of epistemic trust in the musical domain, as it enables certain aspects of early musical enculturation. The same gaze dynamics suggested a lack of mentalisation in some children when approaching music and need for their caregiver's mentalisation, while other children's long bouts of silent, autonomous attention to musical stimuli suggested a much more developed mentalization capacity.

The second and third studies' results also provided evidence for a role of cultural learning in children's attention to music. In discussing the second study, the idea emerged that children might not be simply drawing their attention to recorded music (to which parents were paying attention to), but rather attempting to appraise musical engagement from their parent's perspective in an attempt to understand what was the adult's interest in it and what was their motivation for listening to recorded, non-live music. However, the second study provided no direct evidence in terms of children's gaze, evidence needed in order to better understand whether it was cultural learning that was taking place during interaction with a significant other.

Yet from another angle, the reported gaze dynamics, can be interpreted in terms of cultural learning. Some children consistently directed their gaze at their mothers as a first response to the sound of recorded music, often then imitating the parent's responses. This finding provides preliminary support for the idea that cultural learning in general, and imitative learning in the domain of object-directed action (in this case, culturally-appropriate responses to recorded music) in particular, understood as the child reproducing the mother's actual behaviour in the appropriate functional context, may also take place. On the other hand, other children evidenced considerably lower rates of gaze directed at their parents, and significantly more often directed their cumulative fixation at the music source in an autonomous and attentive manner. As discussed, this more self-sufficient appraisal of recorded music seems to correspond to cultural learning's second step: instructed learning, in which children have— at least partially —internalized their significant other's reactions to music and are thus capable of re-enacting them.

Stemming from the last point, an interesting aspect to discuss is the relationship between cultural learning and attachment. Discussing such a relationship will also allow us to discuss the motivational dimension of this dissertation and to further integrate the different constructs related to socio-cultural interactions so far discussed. In its original formulation (Tomasello et al., 1993), there is no mention whatsoever of a possible role of attachment in cultural learning. Later criticisms reacted to such an omission and portrayed it as a neglect for the emotional and motivational side of cultural learning, especially with regard to adult-child attachment processes (Chisholm & Wescombe, 1994; Marcello, 2006). In Chisholm and Wescombe's (1994) view, the child's capacity for intersubjectivity depends on a crucial motivational or emotional component, such as attachment.

Chisholm and Wescombe's criticisms found echo in subsequent literature stemming from attachment theory. Further research has provided evidence suggesting that mentalisation fosters the child's emotional self-regulation (Slade et al., 2005) and explorative behaviour, in theory enabled by oxytocin and related feelings of autonomy and agency (Luyten & Blatt, 2013).

At the same time, attachment allows caregivers to scaffolds the transmission of cultural knowledge (Fonagy et al., 2007), notably through the ToNP. In the same line, the present work provides further concrete examples of how attachment enables socio-cultural interaction. As already discussed (Chapter 3, section 3.2), children are actively searching for caregiver's affection as well as instances that would provide it in order to achieve a feeling of perceived safety, which constitutes an essential part of attachment. Also as discussed (Chapter 3, section 2.2.2.1), the IDReg is thought to have evolved to convey affective investment. For this reason, the child does not need to learn from an adult how to react to the IDReg but can simply rely on the innate character of animal communication as defined in chapter 2, part I, section 1.1.1. Joint engagement with recorded music, on the other hand, is an instance involving affection that needs to be learned. Accordingly (and fuelled by attachment and by means of mentalisation, epistemic trust and cultural learning), children— actively —inherit from their caregivers the notion that humans cultivate positive affect around recorded music, first through the triadic setting child/caregiver/recorded music (imitative learning), and later directly with recorded music (instructed learning).

Tomasello and collaborators criticize the idea that children must first understand their own intentional states (i.e. understanding what engaging with recorded music is good for) before they may use them to simulate the perspective of others (Tomasello et al., 1993). Basing their judgement in empirical work, the authors categorically refused such a view, instead defending a contrary notion in which children do not understand their own mental states before they understand the mental states of others (Gopnik, 1993). Such a position resonates not just with findings in the last chapter, but also with the main thesis' idea that listening to— and enjoying —recorded music is an activity at first opaque to children (in Sperber's (2010) sense), and that only through musical enculturation does it become natural. This is also consistent with a previous description of the transition that children's musicality undertakes in terms of its social component between the first and third years of life (Forrester, 2009). Forrester notes the manner in which a child initially finds expression and form in closely-synchronized, dyadic interaction, only gradually leading to more self-focused and individuated practices.

2. Revising the thesis

2.1. 'Centripetal' versus 'centrifugal' phenomena

Discussion of the two main findings allows for a first, deeper reassessment of the thinking put forward in chapter 5. On the one hand, parental speech data did not evidence any systematic, direct impact on children's attention to recorded music. On the other hand, the role of interaction— considered, but not foregrounded in the main thesis —was illustrated with a degree of convergence by the three empirical studies. By this token, evidence hereby presented seems to indicate that the choices made in the fifth chapter were not particularly helpful when approaching children's approach to recorded music. Let us decompress this idea. In chapter 5 (section 2.2), I chose to focus on and assess the possible impact of parental use of IDSp on children's listening to recorded music. Such a choice expected to highlight the role of IDSp, while still considering it one element amid other relevant factors such as interaction and other uses of IDSp— notably, the ToNP. In this regard, present data suggests that the relative impact of VAW on children's motivation for engaging with recorded music is substantially less considerable than the mentioned, surrounding factors.

The question then is: why did the thesis of VAW and listening to recorded music as a compensatory measure highlighted an aspect of IDSp that did not prove to be significant, while willingly pushing into the background the ToNP? My explanation is that the thesis failed to correctly ponder 'centripetal' and 'centrifugal' aspects of attachment.

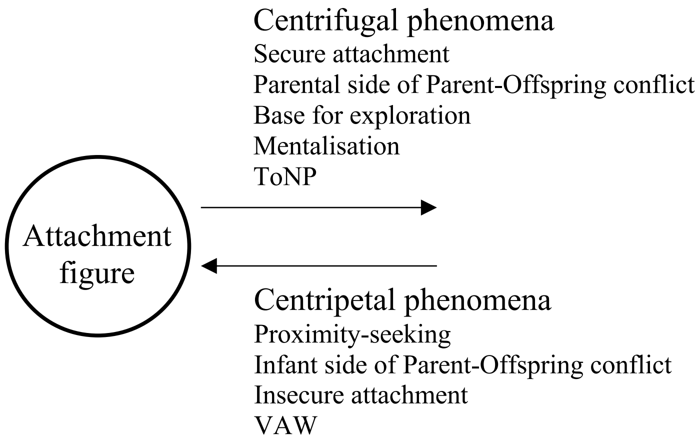


Figure 1. Schematic representation of the centripetal and centrifugal aspects of attachment.

In both Bowlby's (1982) original formulation of attachment and the research on imprinting it was based on, seeking proximity to the attachment figure(s) was a key aspect. For the purpose of illustration, if we consider the attachment figure as the 'centre' of the infant's world, their (human or not) proximity-seeking behaviours can be considered as 'centripetal'— a vector that points towards the centre (Figure 1). Such is the logic behind Lorentz's (1970) proximity-seeking behaviours, many of Bowlby's behavioural systems, the infants' side of Owings and Zeifman's (2004) infant-carer acoustic feedback (chapter 2, section 2.1.2.4), and a large extent of Soltis' (2004) functional analysis of early infant cry (chapter 3, section 2.2.1). Crucially, a centripetal component can be found in parent-offspring (Trivers, 1974) in general, and the idea that IDSi was selected in evolutionary terms precisely as a response to infants' demand for proximity in particular (Falk, 2004: Mehr & Krasnow, 2017). Borrowing Falk's (2004) 'communicating vessels' logic, the main thesis held that toddler's interest in recorded music would increase somewhat proportionally to the unavailability of IDSp, just as prehistoric infants' interest IDSp is thought to have increased proportionally to the impossibility of clinging— and therefore of skin-to-skin contact. In other words, recorded music was regarded as a means for securing a centripetal need, and

IDSp's role was thought to be essentially an affective one. This logic is reflected, for instance, in some of the preliminary (non-operationalised) research questions (e.g. after having experienced the luxury of infant-directed speech and infant-directed singing, why would a child *settle* for a paler alternative such as a recorded track?).

However, and although keeping them in mind, the thesis underestimated other, 'centrifugal' dynamics— following the image of a vector pointing away from the centre (see also Figure 1). Also already in Bowlby's original formulation, the successful development of attachment contemplates children eventually using the main attachment figure as secure bases for the progressive exploration of, and familiarization with ever new elements in the environment. Consequently, for instance, the objects of attachment and dyadic regulation tend to diversify to include other adults and peers (e.g. Doherty & Feeney, 2004; Hazan & Zeifman, 1994). Once again relevant, mentalisation's fostering of the child's emotional self-regulation, explorative behaviour, and feelings of autonomy and agency also work in a centrifugal manner (Luyten & Blatt, 2013; Slade et al., 2005). Closely related, a further critical centrifugal dynamic is the use of IDSp as an ostensive cue at the service of natural pedagogy. As discussed in chapter 3 (section 1.3), and important function of attachment is to provide a supporting framework for the transmission of cultural knowledge (Fonagy et al., 2007)— including the culture's approach to music. Already the first study returned a prevalent use of IDSp as an ostensive cue, and the third provided some evidence for the idea that children at the upper limit of the selected developmental window received more of it than those at the lower limit. Such a use of IDSp differs substantially from the idea of a VAW. First, in this case IDSp participates in a centrifugal dynamic: the exploration of the surrounding world. Second, while in the main thesis the role of IDSp was mainly affective, in the ToNP the affective quality of IDSP is subordinate to (i.e. used as a means for) rather cognitive ends: learning.

One last element to consider potentially against the existence of VAW as initially conceived, are developmental changes in the children's attention or preference for IDSp. As previously mentioned (chapter 5, section 3.3), although experimental evidence of developmental changes in children's preference for IDSp over time is rather limited and does not yet fully converge, it seems hard to deny the existence of such changes (Soderstrom, 2007). Even though not much can be said in empirical terms about this matter— the third study's IDSp data comes from audio recordings and therefore children's attention could not be properly assessed —, a theoretical argument can still be made. In the main thesis, VAW was though to generate conflict because children's— invariable —interest in IDSp triggered a demand. However, following the portion of Hayashi and collaborators' (2001) evidence that has survived replication, if both parental use of IDSp *and* children's interest in it decrease, less or no grounds for conflict should be found. In other words, if both phenomena decrease in parallel and organically, there would be no need for music as a compensatory resource on the children's end. Hayashi (2011) suggests that children's emotional attachment to the exaggerated prosody of IDS wanes in the second phase, an argument not further developed. Following the reasoning at the end of chapter 4, Hayashi's 'exaggerated prosody of IDS' should mainly correspond to IDSp's 'loving tone'. In this regard, data from the first and third studies as well as reasoning in this discussion do not support the notion of a noticeable decline in the use such an acoustic quality. Yet, this does not necessarily mean that children's interest in it might recede. In sum, more empirical data is needed describing exactly what prosodic elements do Hayashi's 'exaggerated prosody of IDS' correspond to.

Adding a further layer to the critical inquiry of my own initial reasoning, it can be said that the mentioned centripetal bias obeys to a certain negative, regressive, or retrograde appraisal of attachment theory. Such an understanding was fuelled by the scope of parent-offspring conflict, and the idea that infants and children resist changes that imply shifts towards their own effort and investment. The bias was also fuelled by some Freudian and Kleinian ideas concerning children and adults' phantasies of returning to symbiotic mother-child states (Maizels, 1985) Bowlby was well aware of and partially agreed with (Bowlby, 1960). Although both arguments share similarities, there is a nuance that separates them and that I believe I previously missed. While the scope of parent-offspring conflict aims to depict a behavioural reality, psychoanalytical (in general) and attachment (in particular) theory contextualize and interpret such behaviours in the frame of normative or desirable psychological development. In other words, while Trivers described the way in which human and non-human infants strive for parental investment, Freud and Klein's concern was to determine to what extent such a behaviour may or may not reflect depressive or neurotic traits (see Maizels, 1985). To put it in yet further terms, even if all infants display 'centripetal' tendencies, individuals vary in terms of the age and extent to which such tendencies persist, and such a persistence may reflect on their attachment and more general mental health.

Thus, after contrasting this dissertation's main thesis with the sum of the collected results, it seems to me that my choice of privileging centripetal phenomena over centrifugal ones turned out to be, to a considerable extent, unwarranted.

The scrutiny of the thesis undertaken in the above paragraphs and the introduced nuances can in fact lead to a new refinement of the thesis, and towards further research. In the thesis presented in chapter 5, I sought for something in recorded music that I thought parents were drawn to providing in lesser quantity through their prosody: positive affect. Instead, data suggests that what parents are doing is fostering the exploration of new elements in the environment. In theory, secure and insecure attachment styles should generate different approaches to new stimuli— such as recorded music. Furthermore, the mentioned fostering should be the product of the participating parent's 'good-enough' parenting style and mentalisation capacities, as well as a certain corresponding security in their children's attachment (see Mikulincer & Shaver, 2007, and further references in chapter 3, section 1.3). Nevertheless, attachment style was not controlled or assessed in any of this work's three studies. Therefore, any further replication or extension of the present studies— especially the third one —should directly assess parental attachment. To this end, the Parental Reflective Functioning Questionnaire (PRFQ) (Luyten et al., 2017) and/or the Adult Attachment Interview (AAI) (Roisman et al., 2007) might be able to confirm this hypothesis. If anywhere, a predominance of centripetal dynamics such as the one postulated in chapter 5 should be found in cases of insecure attachment. For instance, it can be theorized that adults with attachment issues as reflected by the PRFQ or the AAI can be expected to listen to recorded music that reminisces of their attachment figures, in an attempt to compensate for feelings of loss. Something similar could be expected from adults with depressive or neurotic traits.

2.2. Beyond VAW. Factors mediating children's attention to (recorded) music

Beyond the main thesis and its specificities and constraints, data presented in this work allow us to draw a better picture in terms of the development of children's motivation for paying sustained attention to music (in general), and recorded music (in particular). First, the second, empirical half of the present work seems to partially validate the table presented in Chapter 5 (section 3.1, Table 1), where factors mediating infant attention to music as well as the different possible combinations of these factors were arranged in a tentative order of their relative level of appeal.

Let us remember that in the literature, only a partial extent of the factors contemplated in our table were considered in systematic accounts. Perhaps the most relevant example of this is Mehr and Krasnow's (2017) theory on the evolution of IDSi, a relatively recent article that explicitly addresses the matter of factors mediating infant attention. In their theoretical analysis of vocalization and music in terms of honest signals (see chapter 2, part I, section 1.3), Mehr and Krasnow did predict that infant-directed song would generate the strongest attraction, followed by non-infant-directed music, with non-musical vocalizations being be least effective. Although such a ranking both seems theoretically sound and partially matches the evidence hereby presented, it remains incomplete by leaving aside further factors discussed and tested throughout the dissertation, such as the addressee effect (Fernald, 1989), the distance between the child and the caregiver, whether the signal is live or recorded, and the consequent presence or absence of multimodality.

The table in chapter 5 (which for practical reasons we will hereby refer to as the 'old' table) constituted a first effort to incorporate factors so far largely omitted, thus already rendering an unusually comprehensive list of factors to be considered when approaching the subject of infant or child attention to music. Nevertheless, it was (not deliberately) biased by Mehr & Krasnows' thinking, the focus of which was different from the present one. Indeed, the authors had in mind newborns and infants at a certain distance from their mothers, and the relative soothing capacity of auditory signals. For children in such a stage of development— when audition is far more developed than vision —, the source of the musical stimulus is capital, with vocal stimuli definitively being more appealing than others. Thus, the Source factor was placed at the leftmost side of the old table, conveying a role more crucial than that of familiarity or the degree of co-adaptation.

However, such a choice does not fit best the developmental window scrutinized in this dissertation (around 18-24 months of age), in which children can equally see and hear. Results from the second study quite strongly suggested that it was the degree of co-adaptation, and not the source, that predicted attention to music— recorded or not. In other words, nothing seemed to be more appealing for a toddler that somebody's presence and the interaction such a presence implies. The role of interaction was much better illustrated by results from the third study. On the one hand, quantitative and qualitative data converged in suggesting that on the exception of one case study, all assessed toddlers presented some increment in their bout length averages around the eight session, when the piece to the sound of which they had been interacting off-camera was added to on-camera interactions. On the other hand, data also suggested that children eventually perceived the before/after week eight contrast and, because parents were no longer playing an active role around the selected songs, the appeal of the latter tended to fade. These results do not provide strong evidence regarding whether it was evaluative conditioning (EC) or episodic memory (EM) that triggered the children's rather momentary boost in spontaneous sustained attention to recorded music (see Juslin & Vastfjall, 2008, and related discussion in chapter 5, section 2.2). The fact that EM has been associated to tends interpersonal relationships (Baumgartner, 1992) would, in theory, be an argument for its involvement. On the other hand, the fact that the effect tended to wane after week eight despite participating parents being still present in the recorded sessions rather points in the direction of conditioning.

In any case, the sum of these results thus suggests that— in the case of toddlers —co-adaptation is overall more important than the source of the music. These thoughts are reflected in a second, 'new' table (Table 1), aimed to better fit the developmental window that concerned us throughout this dissertation. In it, the degree of co-adaptation has been placed at the leftmost side of the old table, followed by Familiarity, and only then by Source. Contradicting the old table, the live performance of a musical instrument will most likely be more appealing than a recorded song because of the interaction the former generates and despite the appealing timbre of the latter. Accordingly, in the new table such a condition is ranked 5th, as opposed to the old table's 17th position. Notably, despite the described rearrangement, conditions 1-4 in the old table remain unchanged in the new one, and so does condition n.20. In other words, much like in the case of infants IDSi performed live by a familiar person remains the most appealing musical stimulus for toddlers, and a recording of unfamiliar music the least engaging one. The relative appeal of the remaining conditions (6-19) is harder to asses and I no longer dare suggesting a tentative ranking for them. They are therefore left with a question mark in the new table.

Degree of co-adaptation	Familiarity	Source	Register	Type of Vocalization	Condition
Live	Familiar performer	Vocal	ID	Song	1
				Speech	2
			AD	Song	3
				Speech	4
		Instrumental	n/a		5
	Unfamiliar performer	Vocal	ID	Song	?
				Speech	?
			AD	Song	?
				Speech	?
		Instrumental	n/a		?
Recorded	Familiar performer	Vocal	ID	Song	?
				Speech	?
			AD	Song	?
				Speech	?
		Instrumental	n/a	n/a	?
	Unfamiliar performer	Vocal	ID	Song	?
				Speech	?
			AD	Song	?
				Speech	?
		Instrumental	n/a	n/a	20

Table 1. Revised ('new') table of factors mediating infant attention to music of children in the third study. Combinations of these factors are presented in a tentative relative level of attractiveness under the 'Condition' column.

Regarding Mehr and Krasnow's (2017) ranking, while 'infant-directed song' would unambiguously correspond to Condition 1 on the table, 'non-infant-directed music' could correspond to condition 3 and other 6 below it. Something similar applies to non-musical vocalizations. The table thus allows to break one apparently simple notion such as 'non-infant-directed music' down to six possible, more precise alternatives. At a broader level, factor tables such as the ones suggested in this dissertation should be reflected on and rearranged in an adhocratic manner when facing any given developmental window, hopefully assisting a more precise and comprehensive approach to musical and/or speech phenomena. Just as the old table might better fit another developmental window such as infancy and the new one seems to better describe present results, new tables will perhaps contribute to research yet to come.

Finally, even if the data gathered overall did not provide evidence for an impact of infant language acquisition or VAW on attention to recorded music, it still allows us to fill in some of the gaps concerning the timeline presented in chapter 5, section 4.4, concerning the developmental window during which children seem to start paying sustained attention to recorded music. Figure 1 provides a visual summary of the main findings of the present work concerning attention to music and inserts them in an amended version of the developmental timeline. We know that infants as young as four months of age are able to pay sustained attention to IDSi, or briefly react to recorded music, and that certain socially-relevant pieces of music can foster later sustained attention (Mehr, 2016). We also know that as soon as toddlers learn their first words, they can reproduce words or phrases of songs (Hargreaves, 1986) and therefore manage to ask for their favourite tunes, as acknowledged by parents in the first study.

As revealed by descriptive statistics of the second study, at 18 months of age children are still receiving plenty of live ADSi and IDSi (Conditions 3 and 1 in Table 1— some of the most attractive conditions) during daytime and bedtime. Most probably due to such abundance of live singing, children do not seem to be motivated yet to autonomously pay sustained attention to recorded music. Nevertheless, at the same age, the addition of interaction with a caregiver does seem to be motivating enough, as suggested by the second study's Random Forest analysis. Furthermore, our understanding of this phenomenon was deepened by means of our third study, which suggested that such motivation can be explained by attachment and imitative learning, which has started to take place already before 15 months of age. It is worth mentioning that all matters discussed in this paragraph most probably also start taking place before 18 months of age; however, since children as young as 15-months-old were not assessed in the second study, such a suggestion remains speculative.

As suggested by results from the third study, at some point before 23 months of age, and even earlier if supplementary stimulation such as incorporation in music groups take place, children's socio-cultural development imply that the child's mentalisation allows for internalised reactions to music to be more autonomously displayed. Some of these responses to music might represent forms of internalised social cognition (Koelsch 2010) or covert performance (Cross, 2010). Similarly, imitative learning of the caregiver's behaviours when it comes to listening to recorded music is progressively replaced by instructive learning. As a consequence, children are able to autonomously re-enact what they have already learned, which manifests as children paying sustained attention to the recorded music they like. Data from the third study also confirms that such a capacity does not necessarily imply a corresponding absence of IDSi. Finally, a gap is still left between children 27 months-old and their 38-46 months-old counterparts. Lamont's (2008) study indicates that by that age, the amount of IDSi that children receive during daytime has greatly decreased.

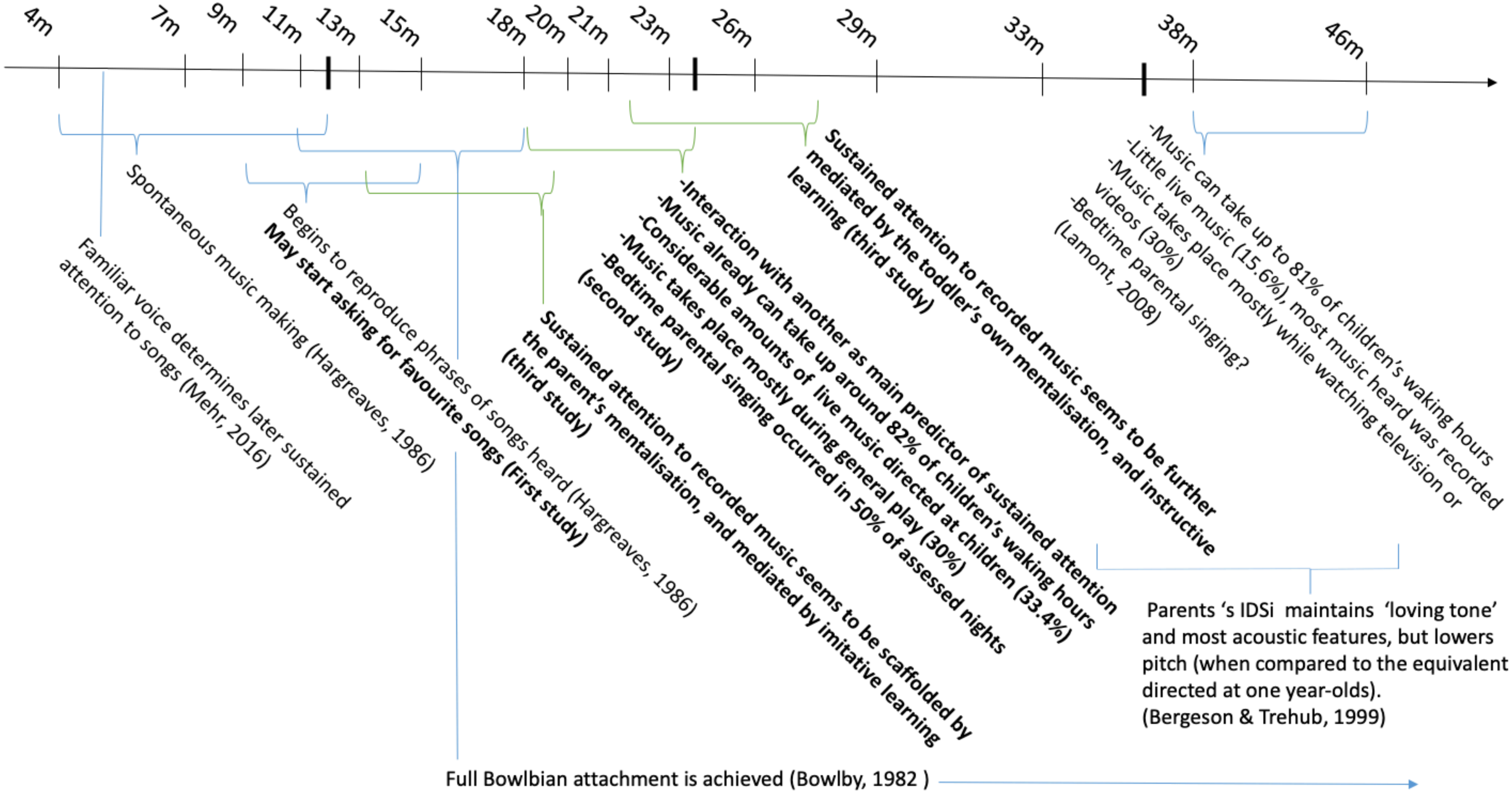


Figure 1. Amended timeline summarizing the development of infant attention to music. Information derived from the studies undertaken for this thesis (refer to the respective chapter's discussion for further details) is presented in bold font, and the braces that relate such finding to the timeline are green (instead of light blue).

2.3. A note on the pros and cons of an analytical approach

As hinted in the first chapter, this dissertation sought to focus on autonomous listening to recorded music, and a progressive effort was made towards distinguishing the targeted phenomenon from concomitant factors. Closely-related phenomena such as interpersonal coordination (chapter, section 1.4), typical musical activities (chapter 5, section 3.5) and IDSi (chapter 3, section 2.2.2 and chapter 4, section 1) were duly acknowledged, yet an effort was made to explicitly detach from them (chapter 5, section 2.2). These choices are best reflected in the third study, were a situation was devised so as to generate the exact desired situation: children faced to recorded music and being able to react to it spontaneously. The mentioned setting was thus the result of an analytical exercise, and deliberately stripped of undesired elements that could prime or bias the toddlers' reactions.

Detaching from the typical ways in which toddlers are exposed to music has both desirable and undesirable consequences. While gaining precision, the approach loses ecological validity. In other words, the approach is not only analytical but also eccentric to some degree, and the reader might legitimately ask themself what is the point— if any —in creating situations that are not to be found as such in real life. In this section, I will thus aim to explain why and how the dissertation proves a relevant, if costly point. In order to do so, I will compare the present dissertation to an already discussed article by Singh et al. (2002).

As discussed in chapters 4 (section 3.2) and 6 (section 4), Singh and collaborators systematically manipulated affect and speech register, generating experimental situations where 'neutral IDSp' and 'sad IDSp' utterances were directed at children. The authors themselves noted that such an analytical exercise (dissecting elements and recombining them) is uncustomary in the literature, and I have already stated that it entails a degree of artificiality. Notwithstanding, such artificiality succeeded in illustrating a pervasive confound between two factors: affect and register, without denying that some combinations are more natural and common than others. Accordingly, the first study's results echoed those of Singh and collaborators', with instances of 'IDSp happy' being more frequently acknowledged and more easily described by participating parents than 'IDSp fear' or 'IDSp anger'.

I believe something similar can be said about the present dissertation. I systematically manipulated interaction and music, and generated experimental situations where parents were present yet did not prompt any particular musical activity nor sung. Precisely, the situation illustrated to what extent sustained attention to recorded music depended on interaction with a significant other; how the former waned when the latter was missing. Far from denying the role of interaction discussed in chapter 3 (section 1.4), the third study confirms and strengthens it by testifying the effects of its absence. By this token, 'interaction' is the answer to some of my preliminary research questions: Can some kind of continuity be observed between children's initial motivation to engage with IDSp or IDSi, and their later motivation for engaging with recorded Western music? Are there any extra-musical phenomena that could help explain such a continuity? In other words, rather than caregiver's everyday use of IDSp, it is interaction in the form of mutual synchronisation, joint singing, or the simple confirmation of elements in the music that might remain opaque to the toddler (e.g. appropriate emotional response, actions, etc.) that motivate children to delve into it.

3. Further criticisms and future directions

The present work was, to an extent, undermined by being over-ambitious. The main thesis emerged as an attempt to deal with the problem that there has been research focusing on the musicality of the IDReg on the one hand, and investigation addressing Western children's engagement with recorded music on the other, but the relationship between these two research topics has been implicitly accepted and not explicitly scrutinised. An attempt was made to present the problem first in simple terms (chapter 1) and, after a detailed exposition of the involved literature (chapters 2-4), a second time, in depth (chapter 5). As much as an effort could have been made— and it was made —to present the problem in the clearest possible manner, such an effort could not take away the problem's considerable level of complexity. Furthermore, the thesis' intended contribution to the problem consisted of adding a new phenomenon in an attempt fill the gap between the research topics referred to: vocal-affective weaning.

The addition of VAW was ambitious and risky. It was ambitious because it meant adding, on top of an already complex problem involving the interrelation of two research topics, a new concept that aimed to represent in turn an unresolved question: how, when and why do parents stop enacting the affectionate disinhibition that characterizes IDSp? The addition of VAW was also risky given that it meant heavily relying on the— to date —scarce literature that has assessed changes in parental use of the IDReg, the diminishing curve of the latter being largely unknown, thus forcing us to work with a developmental timeline that entails considerable gaps and only assesses proxies (partial prosodic aspects) of IDSp. The second reason is as important as the first: as its name suggests, VAW fundamentally concerns a decrease in the baseline level of affection and an increase of negative emotions such as anger in the vocal prosody of the utterances that parent's direct to their children as they grow older. Although IDSp indeed comprises a high baseline level of affection, it was never guaranteed that the changes in IDSp's proxies that tended towards ADSp as reported in the literature would be accompanied by an actual corresponding decrease in the baseline level of affection or increase of anger or frustration. In other words, in the main thesis, virtually equating the diminishing curve of the use of the IDReg to VAW was a rather risky assumption.

The risk's appeal was thus the chance of, by means of a few empirical studies, successfully laying the foundations of a theoretically-sound and empirically-supported connection between the musicality of the IDReg and investigation addressing Western children's engagement with recorded music. If the diminishing curve of the use of the IDReg effectively corresponded to VAW, then the changes in parental prosody reported by Ko (2012) and Farran et al. (2016) (see chapter 5, section 3.2.2) would be accompanied by a corresponding decrease in the baseline levels of affection and negative emotions conveyed through vocal prosody.

It was therefore surprising to find no replication of the changes in parental prosody reported either by Farran nor by Ko, even though the third study directly measured parental prosody as Farran et al. did. Indeed. The same can be said regarding speech rate— no quantal changes were found, although I used speech rate just like Ko (2012) did —and all of the f0 measures directly assessed in this work's final study. In addition, no increases in instances of conflict between mothers and children were found, bearing in mind that conflict could have evidenced a progressive expression of negative emotions on the part of the mothers.

Notwithstanding the criticisms raised above, the problem laid out in the first half of this dissertation remains arguably relevant, as much of the relationship between the IDReg and music remains unknown. On the other hand, and although the possibilities for investigating the conception and validity of VAW have not been yet exhausted, theoretical reasons for searching for an incidence of the IDReg on Western children's engagement with recorded music remain less compelling. Instead, future research could confirm and extend results from the third study that suggest changes through time in the use of IDSp as an ostensive cue.

This in turn leads to the larger topic of what happens with the use of IDSp as children grow older. In the same manner as there is no available evidence describing the temporal unfolding of vocal ostensive cues, little is known about IDSp after infancy. As the result of this thesis and its discussion so far, I could postulate at least two scenarios, the first being that VAW exists as described in chapter 5, but it was somehow not properly measured in this dissertation. Future attempts to measure its temporal unfolding should therefore explicitly assess the presence and degree of a 'loving tone' in parental prosody. I find this possibility less likely. As an alternative, it can be conceived that VAW is not a sound concept. I find more arguments in favour of this possibility. First, as above discussed (section 2.1), the concept unwarrantedly overestimates centripetal over centrifugal attachment dynamics. A second reason is that data presented in this dissertation provide some— humble —support for the idea that the 'loving tone' does not really wane along with other prosodic changes such a f0- or time-related measures. After all, affection is expressed to children until much, much later than toddlerhood. Stemming from these two reasons, a third one emerges: present data suggests that even if the affective functions of IDSp be impeded by normativity or other demands, more cognition-oriented uses as described in the ToNP would still be present— and in fact are suggested to increase. Fourth, it might be that parents become more selective in the instances of its use. Finally, the temporal unfolding of VAW might span several years without quantal changes, being so slow and subtle that it would largely elude any attempt at measurement.

In any case, a next study can be outlined that could favour one or more of these possibilities while rather discarding others. It would consist of a partial replication of the third study, with a sample equal to or larger than the second study (25-30 mother infant-dyads) so as to test statistically what has so far been only suggested. To this end and bearing in mind all the hardship that recruitment entailed in this dissertation's studies, the hiring of a local, female research assistant should be considered, as she would most probably be more readily trusted than myself. The Language Use Inventory and the One-way Remote Video Recording Technique seem to have successfully assessed children's pragmatic development and attention to music respectively and thus should be employed. However, any manipulation of the interaction between parents and children should be considered and, if necessary, carefully executed. In the assessment of parental prosody, changes should also be contemplated. Although f0- and time-related measures could be kept, vocal assessment should focus first and foremost on the 'loving tone' by means of the qualitative appraisal of independent judges, as reported by Trainor (1996). A suitable, perhaps complementary alternative would be having the same independent judges rating the 'loving tone' of IDSp samples of the targeted developmental windows, available in databanks like the CHILDES database (https://childes.talkbank.org). Such a study could focus on assessing whether the 'loving' quality decreases in parallel to changes in f0 and/or speech rate. This comparatively-less comprehensive approach, precisely because of being simpler and more succinct, could precede the potential study outlined above by serving as a pilot or prequel.

As a final element of the main thesis, the case was also contemplated that a negative correlation between VAW and interest in music would become stronger (quantal changes should be found) with the arrival of puberty, when the former child commences the slow process of shifting their main attachment figure. In this period of transition, young adolescents seek to attach to potential partners, as in peers with whom strong affective or romantic attachments can be formed. The resulting affective instability would, in all likelihood, be connected with the increase in music consumption as well as the intensity of that consumption that characterize puberty and adolescence (Bogt, Delsing, Zalk, Christensen, & Meeus, 2011; Cohrdes & Kopiez, 2015; Nuttall, 2008).

As mentioned, VAW might in fact be a chimera because parents may progressively change most prosodic parameters that characterize IDSp such as f0 mean, f0 range and speech rate, while never fully changing their use of the 'loving tune', or doing so to an insignificant degree. After all, parents remain affectionate, as demonstrated by participants in the first study. It could well be the case that parents become more selective in the instances of its use of IDSp and its 'loving tone. As discussed in chapter 3, section I.2.3, a continuity between infant and adult attachment can be acknowledged in the circumstances that lead an adult's attachment behaviour to become more readily prompted (e.g. sickness, danger, or calamity). Accordingly, as the child grows up and reaches ever-growing levels of emotional independence, these instances would become more and more rare. For instance, although during puberty individuals would normally overtly avoid physical proximity or overt manifestations of love towards or from their attachment figures, emotionally-challenging circumstances might nevertheless induce in them proximity-seeking behaviours and pursue perceived safety. In such circumstances mothers, although normally refraining from doing so may resort to vocalizations with an overt 'loving tone'. The difficulty would then lie in identifying these circumstances and empirically assessing them.

It could also be the case that the prosodic variations that VAW entails take place so slowly and smoothly between infancy and puberty that it becomes virtually impossible to assess the change by means other than a decade-long longitudinal study. Furthermore, such slow and smooth variations in the 'loving tone' could be so to a degree that straightforwardly eludes any judge's qualitative appraisal. In any of the scenarios contemplated in this paragraph it would, in effect, be simply the case that parental use of IDSp recedes while children's attentional capacities augment, with no direct connection between the two phenomena.

A final aspect of this dissertation that could lead to further research is the ethnomusicological dimension. As discussed in chapter 1, our focus on a Western conception of music and musical enculturation left aside a wealth of existing alternatives around the globe. Nevertheless, most of the remaining theoretical aspects of the thesis did not suffer such a cultural narrowing. Indeed, elements such as attachment, parent-offspring conflict, the IDReg, and the language/music dichotomy remain universal. Therefore, the interplay between such elements and other, non-Western form of music and enculturation could be equally explored.

4. Conclusions

The present dissertation sought to theorize, empirically-assess, and describe how changes in the use of IDReg may provide motivation for cultural practices such as listening to recorded music, the latter being a central form of Western musical enculturation. Its main thesis predicted that it would be changes in parental use of the IDSp that would— at least partially —foster changes in children's engagement with recorded music. However, no substantial evidence from the empirical studies support such a prediction.

IDSp seems indeed to be involved in children's initial and progressive engagement with music; not as a compensation for any loss of affection, but as scaffolding used in exploring a still opaque cultural product such as a record. Data suggest a shift in the prevalence of different functions if IDSp, rather than a decrease in its overall use. In this sense, the dissertation contributes some humble insights regarding a potential, relative importance of affectivity (as in VAW) and cognition (as in ToNP) as parallel, coexisting governing principles that exert an influence on developmental changes concerning parental use of IDSp.

Alternatively, data seem to confirm previous existing literature in suggesting that interaction is a phenomenon that fosters toddlers' attention to recorded music. Furthermore, interaction seems to serve as an intermediary scaffolding resource between IDSi— likely the most appealing musical stimulus —and autonomous (solitary) listening— a much less appealing one. In other words, there are many more signals that can be shared between a caregiver and a toddler, than vocal ones alone. By this token, although not as initially hypothesised, the present dissertation does contribute elements to an explanatory chain linking the IDReg and Western forms of music engagement.

Closely related to the above paragraph, forms of socio-affective and socio-cultural interaction such as mentalisation and cultural learning seem to enable toddlers' sustained attention to recorded music, apparently without the necessary intervention of quantal changes in parental use of the IDSp. The fact that results largely contradict the main thesis— an outcome entirely possible in empirical research —can also be largely explained through my unwarranted privileging of centripetal over centrifugal attachment dynamics. In other words, I formulated the thesis in somewhat negative, retrograde, or regressive terms, thus leaving no room for the thrill of exploring a whole new world. Indeed, IDSp seems to be (more and more) about learning, as attachment is hopefully (more and more) about exploring. The present work nevertheless provides advances our understanding of how children gradually depend less on actual interaction, and interiorize such an interaction into solitary, autonomous listening.

Beyond empirical results, the theoretical understanding of the IDReg and related phenomena presented in the first half of this work remains novel in its depth. It synthesizes for the first time a rather vast amount of literature concerning its foundations in communication and attachment theories. Such a theoretical grounding would most probably prove useful for future work in the field, be it theoretical or empirical.

Also beyond the conceptualization of VAW, its hypothetic role in musical enculturation, and the extent to which the main thesis could be tested in this dissertation, I would still contest that the problem described and addressed in this dissertation remains a relevant one. A continuity between the IDReg and subsequent engagement with Western recorded music remains largely taken for granted yet not explained. If such a pervasive— and compelling —view, ever present in the most important contemporary handbooks of music psychology, is to be taken seriously, not just the present study, but substantially more efforts either along its lines or from novel standpoints, will need to be undertaken.

REFERENCES

Ahumada-Galleguillos, P., Fernández, M., Marin, G. J., Letelier, J. C., & Mpodozis, J. (2015). Anatomical organization of the visual dorsal ventricular ridge in the chick (Gallus gallus): layers and columns in the avian pallium. *Journal of Comparative Neurology, 523*(17), 2618-2636.

Aiello, L. C., & Dunbar, R. I. M. (1993). Neocortex Size, Group Size, and the Evolution of Language. *Current Anthropology, 34*(2), 184-193. doi:10.1086/204160

Ainsworth, M., & Wittig, B. (1969). Attachment, exploration, and separation: illustrated by the behavior of one-year-olds in a strange situation. *Determinants of infant behaviour, 4*, 113-136.

Allen, J. P., & Land, D. (1999). Attachment in adolescence. *Handbook of attachment: Theory, research, and clinical applications*, 319-335.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behav. Brain Sci., 28*(2), 105-124. doi:10.1017/S0140525X05000038

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology, 18*(5), R202-R203. doi:https://doi.org/10.1016/j.cub.2008.01.040

Arom, S. (1991). *African Polyphony and Polyrhythm*. Cambridge: Cambridge University Press.

Atkinson, C. M. (2008). *The critical nexus: Tone-system, mode, and notation in early medieval music*. Oxford University Press.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American journal of psychology*, 147-166.

Bail, A., Morini, G., & Newman, R. S. (2015). Look at the gato! Code-switching in speech to toddlers. *Journal of child language, 42*(05), 1073-1101.

Barrett, M. S. (2010). Musical narratives: A study of a young child's identity work in and through music-making. *Psychology of Music, 39*(4), 403-423.

Baumgartner, H. (1992). Remembrance of Things Past: Music, Autobiographical Memory, and Emotion. *Advances in Consumer Research*, 19(1).

Beebe, B., Jaffe, J., Feldstein, S., Mays, K., & Alson, D. (1985). Interpersonal timing: The application of an adult dialogue model to mother-infant vocal and kinesic interactions. *Social Perception in Infants*, 217–247.

Bergeson, T. R., & Trehub, S. E. (1999). Mothers' singing to infants and preschool children. *Infant Behavior and Development, 22*(1), 51-64.

Bertau, M.-C. (2007). On the notion of voice: An exploration from a psycholinguistic perspective with developmental implications. *International Journal for Dialogical Science, 2*(1), 133-161.

Biro, S., Alink, L. R. A., Huffmeijer, R., Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (2017). Attachment quality is related to the synchrony of mother and infant monitoring patterns. *Attachment & Human Development. Volume 19:Issue 3 (2017, June)*.

Blacking, J. (1995). *Music, Culture and Experience*. London: University of Chicago Press.

Blurton Jones, N. (1972). Comparative aspects of mother-child contact. In N. Blurton Jones (Ed.), *Ethological studies of child behaviour* (pp. 305-328). London: Cambridge University Press.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. (Version 6.1.03). Retrieved from http://www.praat.org/

Bogin, B. (1999). Evolutionary perspective on human growth. *Annual Review of Anthropology*, 109-153.

Bogt, T. F. M. t., Delsing, M. J. M. H., Zalk, M. v., Christensen, P. G., & Meeus, W. H. J. (2011). Intergenerational Continuity of Taste: Parental and Adolescent Music Preferences. *Social Forces, 90*, 297-319.

Bolinger, D. (1986). *Intonation and its parts: melody in spoken English*. Stanford, Calif: Stanford, Calif : Stanford University Press, 1986.

Bombar, M. L., & Littig, L. W. (1996). Babytalk as a communication of intimate attachment: An initial study in adult romances and friendships. *Personal Relationships, 3*(2), 137-158.

Bourke, L. & Adams, A.M. (2012). Is it differences in language skills and working memory that account for girls being better at writing than boys? *Journal of Writing Research*, *3*(3), 249–277. https://doi.org/10.17239/jowr-2012.03.03.5

Bowlby, J. (1958). The nature of the child's tie to his mother. *The International journal of psycho-analysis, 39*, 350.

Bowlby, J. (1960). Separation anxiety: A critical review of the literature. *Journal of Child Psychology and Psychiatry*, 1(4), 251–269.

Bowlby, J. (1969). Attachment: Attachment and loss (vol. 1). *London: Hogarth*.

Bowlby, J. (1982). *Attachment: Attachment and loss* (2nd ed. Vol. 1). London: Basic Books.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage Publications.

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science, 11*(6), 853-861.

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in psychology, 3*.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences, 113*(6), 1666-1671.

Bretherton, I. (2003). Mary Ainsworth: Insightful observer and courageous theoretician. In *Portraits of Pioneers in Psychology*. American Psychological Association.

Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye... attention capture by infant faces.

Brown, S. (2000). The “musilanguage” model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music*. Cambridge, MA.: MIT Press.

Brown, S. (2001). Are music and language homologues? *Annals-new york Academy of sciences, 930*, 372-374.

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture, 8*(1), 135-148.

Bråten, S. (1988). Dialogic mind: The infant and the adult in protoconversation. In *Nature, Cognition and System I* (pp. 187-205): Springer.

Buck, S. (2011). The Evolutionary history of the modern birth mechanism: looking at skeletal and cultural adaptations. *Totem: The University of Western Ontario Journal of Anthropology, 19*(1).

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 18*(5), 175-186.

Burling, R., Armstrong, D. F., Blount, B. G., Callaghan, C. A., Foster, M. L., King, B. J., . . . Wallace, R. (1993). Primate Calls, Human Language, and Nonverbal Communication [and Comments and Reply]. *Current Anthropology, 34*(1), 25-53.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in cognitive sciences, 12*(5), 187-192.

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, *31*(3), 1327–1342. https://doi.org/10.1016/j.neuroimage.2006.01.036

Campbell, L. (2001). The history of linguistics. In M. Aronoff & J. Rees-Miller (Eds.), *The handbook of linguistics* (pp. 81-104): Blackwell.

Campbell, P. S. (1998). The Musical Cultures of Children. *Research Studies in Music Education*, *11*(1), 42–51. https://doi.org/10.1177/1321103X9801100105

Caporael, L. R. (1981). The paralanguage of caregiving: Baby talk to the institutionalized aged. *Journal of personality and social psychology, 40*(5), 876.

Caporael, L. R., & Culbertson, G. H. (1986). Verbal response modes of baby talk and other speech at institutions for the aged. *Language & Communication, 6*(1-2), 99-112.

Chang, R., & Garcia, J. (2011). *Behind closed doors: On the use of loverese and pet-names between romantic partners.* Paper presented at the Poster session presented at the Annual Conferences of the Northeast Evolutionary Society, Binghamton, NY.

Chang, R. S. (2013). Human attachment vocalizations and the expanding notion of nurture. *Evolution's empress: Darwinian perspectives on the nature of women*, 168-187.

Cheney, D. L., & Seyfarth, R. M. (2010). Primate communication and human language: continuities and discontinuities. In P. M. Kappeler & J. B. Silk (Eds.), *Mind the Gap* (pp. 283-298). Heidelberg: Springer.

Chisholm, J. S., & Wescombe, N. (1994). Evolution, attachment, and cultural learning. *Behav Brain Sci, 17*(4), 778-779. doi:10.1017/S0140525X0003716X

Chomsky, N. (1988a). *Language and problems of knowledge: The Managua lectures* (Vol. 16): MIT press.

Chomsky, N. (1988b). *Prospects for the study of language and mind.* Paper presented at the conference" The Chomskyan Turn," Jerusalem.

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*(6), 1003–1011. https://doi.org/10.1111/desc.12193

Clayton, M., Sager, R., & Will, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. *ESEM Counterpoint*, *1*, 1-45.

Coan, J. A. (2008). Toward a neuroscience of attachment. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2 ed., pp. 241–265). New York: Guilford.

Coan, J. A. (2010). Adult attachment and the brain. *Journal of Social and Personal Relationships, 27*(2), 210-217.

Cohrdes, C., & Kopiez, R. (2015). Optimal distinctiveness and adolescent music appreciation: Development of music- and image-related typicality scales. *Psychology of Music, 43*(4), 578-595. doi:10.1177/0305735614520851

Collias, N. E. (1960). An ecological and functional classification of animal sounds. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal sounds and communication* (Vol. 7). Washington, D.C.: American Institute of Biological Sciences.

Collias, N. E. (2000). Vocal signals of the Village Weaver: a spectrographic key and the communication code. *The Condor, 102*(1), 60-80.

Collins, N. L. (1996). Working Models of Attachment: Implications for Explanation, Emotion, and Behavior. *Journal of Personality and Social Psychology*, *71*(4), 810–832. https://doi.org/10.1037/0022-3514.71.4.810

Colombo, J. (1985). Spectral complexity and infant attention. *The Journal of genetic psychology, 146*(4), 519-525.

Cooper, R. P., & Aslin, R. N. (1994). Developmental Differences in Infant Attention to the Spectral Properties of Infant-directed Speech. *Child Development, 65*(6), 1663-1677.

Corbeil, M., Trehub, S. E., & Peretz, I. (2015). Singing delays the onset of infant distress. *Infancy*.

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: evidence from behavioral and electrophysiological methods. *Developmental science, 17*(1), 142-158.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., Liddle, B., Hearn, A., Vittorini, L., & De Rosnay, M. (2009). Young Children's Trust in Their Mother's Claims: Longitudinal Links With Attachment Security in Infancy. *Child Development*, *80*(3), 750–761. https://doi.org/10.1111/j.1467-8624.2009.01295.x

Costa-Giomi, E., & Ilari, B. (2014). Infants' Preferential Attention to Sung and Spoken Stimuli. *Journal of Research in Music Education*, *62*(2), 188–194. https://doi.org/10.1177/0022429414530564

Cross, I. (2007). Music and cognitive evolution. In I. Dunbar & L. Barrett (Eds.), *Handbook of evolutionary psychology* (pp. 649-667). Oxford: OUP.

Cross, I. (2010). Listening as covert performance. *Journal of the Royal Musical Association, 135*(S1), 67-77.

Cross, I. (2013). "Does not compute"? Music as real-time communicative interaction. *AI & society, 28*(4), 415-430.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music, 42*(6), 809-819.

Cross, I., & Woodruff, G. E. (2009). Music as a communicative medium. In R. Botha & C. Knight (Eds.), *The prehistory of language* (Vol. 11, pp. 77): Oxford University Press.

Crowe, H. P., & Zeskind, P. S. (1992). Psychophysiological and perceptual responses to infant cries varying in pitch: Comparison of adults with low and high scores on the child abuse potential inventory. *Child Abuse & Neglect, 16*(1), 19-29. doi:https://doi.org/10.1016/0145-2134(92)90005-C

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In *Processes of Change in Brain and Cognitive Development: Vol. XXI Attention and Performance* (pp. 249–274). Oxford University Press.

Csibra, Gergely, & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, Gergely, & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B*, *366*(1567), 1149–1157. https://doi.org/10.1098/rstb.2010.0319

Custodero, L. A. (2002). Seeking Challenge, Finding Skill: Flow Experience and Music Education. *Arts Education Policy Review*, *103*(3), 3–9. https://doi.org/10.1080/10632910209600288

Custodero, L. A. (2006). Singing Practices in 10 Families with Young Children. *Journal of Research in Music Education*, *54*(1), 37–56. https://doi.org/10.1177/002242940605400104

Custodero, L. A., & Johnson-Green, E. A. (2003). Passing the Cultural Torch: Musical Experience and Musical Parenting of Infants. *Journal of Research in Music Education*, *51*(2), 102–114. https://doi.org/10.2307/3345844

Custodero, L. A., Rebello Britto, P., & Brooks-Gunn, J. (2003). Musical lives: A collective portrait of American parents and their young children. *Journal of Applied Developmental Psychology*, *24*(5), 553–572. https://doi.org/10.1016/j.appdev.2003.08.005

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*(3), B1-B10.

Davies, S. (1997). John Cage's 4′33″ : Is it music? *Australasian Journal of Philosophy, 75*(4), 448-462. doi:10.1080/00048409712348031

Dawkins, R., Krebs, J. R., Maynard, S. J., & Holliday, R. (1979). Arms races between and within species. *Proc. R. Soc. Lond. B, 205*.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative Learning of Likes and Dislikes: A Review of 25 Years of Research on Human Evaluative Conditioning. *Psychological Bulletin*, *127*(6), 853–869. https://doi.org/10.1037/0033-2909.127.6.853

de L'etoile, S. K. (2006). Infant behavioral responses to infant-directed singing and other maternal interactions. *Infant Behavior and Development, 29*(3), 456-470. doi:10.1016/j.infbeh.2006.03.002

De Waal, F. B. (2005). A century of getting to know the chimpanzee. *Nature, 437*(7055), 56-59.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. London: Penguin.

DeCasper, A. J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant behavior and development, 17*(2), 159-164.

Dessalles, J.-L. (1998). Altruism, status and the origin of relevance. *Approaches to the Evolution of Language*, 130-147.

Dillard, J. P., Solomon, D. H., & Palmer, M. T. (1999). Structuring the concept of relational communication. *Communication Monographs, 66*(1), 49-65. doi:10.1080/03637759909376462

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. *The origins of music*, 389-410.

Dissanayake, E. (2001). Becoming Homo Aestheticus: Sources of aesthetic imagination in mother-infant interactions. *Substance, 30*(1), 85-103.

Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Musicae Scientiae, 12*(1 suppl), 169-195.

Doherty, N. A., & Feeney, J. A. (2004). The composition of attachment networks throughout the adult years. *Personal Relationships*, *11*(4), 469–488. https://doi.org/10.1111/j.1475-6811.2004.00093.x

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution, 22*(6), 469-493. doi:https://doi.org/10.1016/0047-2484(92)90081-J

Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Dunbar, R. I. M. (2004a). Gossip in evolutionary perspective. *Review of General Psychology, 8*(2), 100-110. doi:10.1037/1089-2680.8.2.100

Dunbar, R. I. M. (2004b). Language, Music, and Laughter in Evolutionary Perspective. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

Edwards, J. (2011). The use of music therapy to promote attachment between parents and infants. *The Arts in Psychotherapy*, *38*(3), 190–195.

Ein-Dor, T., & Doron, G. (2015). Psychopathology and attachment. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and research: New directions and emerging themes* (pp. 346-373). London: The Guilford Press.

ELAN (Version Version 5.0.0-beta). (2019). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

Emery, N. J., Seed, A. M., Von Bayern, A. M., & Clayton, N. S. (2007). Cognitive adaptations of social bonding in birds. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362*(1480), 489-505.

Ensink, K., & Mayes, L. C. (2010). The Development of Mentalisation in Children From a Theory of Mind Perspective. *Psychoanalytic Inquiry: How Does a Person Know What's Going on in the Mind of Another?*, *30*(4), 301–337. https://doi.org/10.1080/07351690903206504

Falk, D. (1998). Hominid Brain Evolution: Looks Can Be Deceiving. *Science, 280*(5370), 1714-1714. doi:10.1126/science.280.5370.1714

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences, 27*(04), 491-503.

Falk, D. (2004). Prelinguistic evolution in hominin mothers and babies: For cryin'out loud! *Behavioral and brain sciences, 27*(4), 461-462.

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1491.

Farabaugh, S. M., & Dooling, R. J. (1996). Acoustic communication in parrots: laboratory and field studies of budgerigars, Melopsittacus undulatus. *Ecology and evolution of acoustic communication in birds*, 97-117.

Faraj, A. (1988). Language acquisition and development in infants. *Early Child Development and Care*, *39*(Oct 88), 21–31.

Farran, L. K., Lee, C.-C., Yoo, H., & Oller, D. K. (2016). Cross-Cultural Register Differences in Infant-Directed Speech: An Initial Study. *PLoS One, 11*(3), e0151518. doi:10.1371/journal.pone.0151518

Fawcett, C., & Tunçgenç, B. (2017). Infants’ use of movement synchrony to infer social affiliation in others. *Journal of Experimental Child Psychology (Print)*, *160*(160), 127–136. https://doi.org/10.1016/j.jecp.2017.03.014

Fearon, P., Shmueli-Goetz, Y., Viding, E., Fonagy, P., & Plomin, R. (2014). Genetic and environmental influences on adolescent attachment. *Journal of Child Psychology and Psychiatry*, *55*, 1033–41.

Feldman, R., & Eidelman, A. I. (2004). Parent-infant synchrony and the social-emotional development of triplets. *Dev Psychol*, *40*(6), 1133–1147. https://doi.org/10.1037/0012-1649.40.6.1133

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). MacArthur-Bates Communicative Development Inventories user's guide and technical manual, 2d ed. *Reference and Research Book News, 22*(3).

Ferguson, C. A. (1977). Baby talk as a simplified register. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to Children* (pp. 209-235). Cambridge: Cambridge University Press.

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*, 1497-1510.

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of child development, 8*, 43-80.

Fischer, J., Metz, M., Cheney, D. L., & Seyfarth, R. M. (2001). Baboon responses to graded bark variants. *Animal Behaviour, 61*(5), 925-931. doi:https://doi.org/10.1006/anbe.2000.1687

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences, 4*(7), 258-267.

Fitch, W. T. (2004). Kin Selection and "Mother Tongues": A Neglected Component in Language Evolution. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

Fonagy, P. (2010). *Attachment theory and psychoanalysis*. London: Karnac.

Fonagy, P., & Campbell, C. (2015). Bad Blood Revisited: Attachment and Psychoanalysis, 2015. *British Journal of Psychotherapy. Volume 31:Number 2 (2015, May)*.

Fonagy, P., Gergely, G., & Target, M. (2007). The parent–infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry*, *48*(3–4), 288–328. https://doi.org/10.1111/j.1469-7610.2007.01727.x

Fonagy, P., Luyten, P., & Allison, E. (2015). Epistemic Petrification and the Restoration of Epistemic Trust: A New Conceptualization of Borderline Personality Disorder and Its Psychosocial Treatment. *Journal of Personality Disorders*, *29*(5), 575–609. https://doi.org/10.1521/pedi.2015.29.5.575

Fonagy, P., Steele, H., & Steele, M. (1991). Maternal Representations of Attachment during Pregnancy Predict the Organization of Infant-Mother Attachment at One Year of Age. *Child Development*, *62*(5), 891–905. https://doi.org/10.1111/j.1467-8624.1991.tb01578.x

Ford, J. K. (1991). Vocal traditions among resident killer whales (Orcinus orca) in coastal waters of British Columbia. *Canadian journal of zoology, 69*(6), 1454-1483.

Forrester, M. A. (2009). Emerging musicality during the pre-school years: A case study of one child. *Psychology of Music*, *38*(2), 131–158. https://doi.org/10.1177/0305735609339452

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 177-206.

Fraley, R. C. (2002). Attachment Stability From Infancy to Adulthood: Meta-Analysis and Dynamic Modeling of Developmental Mechanisms. *Personality and Social Psychology Review*, *6*(2), 123–151. https://doi.org/10.1207/S15327957PSPR0602_03

Fraley, R. C., Vicary, A. M., Brumbaugh, C. C., & Roisman, G. I. (2011). Patterns of Stability in Adult Attachment: An Empirical Test of Two Models of Continuity and Change. *Journal of Personality and Social Psychology*, *101*(5), 974–992. https://doi.org/10.1037/a0024150

Frodi, A. (1985). When Empathy Fails. In L. B.M. & Z. B. C.F. (Eds.), *Infant Crying*. Boston, MA: Springer.

Gabbard, C. (2014). Embodied cognition in children: Developing mental representations for action. In S. Robson & S. F. Quinn (Eds.), *The Routledge international handbook of young children's thinking and understanding* (pp. 229–237). Routledge.

Gallacher, L. (2005). 'The terrible twos': Gaining control in the nursery? *Children's Geographies*, *3*(2), 243–264. https://doi.org/10.1080/14733280500161677

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and brain sciences, 23*(04), 573-587.

Gardner, R. A., & Gardner, B. T. (1969). Teaching Sign Language to a Chimpanzee. *Science, 165*(3894), 664-672. doi:10.1126/science.165.3894.664

Gillath, O. (2015). The neuroscience of attachment: Using new methods to answer old (and new) questions. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment Theory and*

Research. New Directions and Emerging Themes (pp. 39-67). London: The Guilford Press.

Glocker, M. L., Langleben, D. D., Ruparel, K., Loughead, J. W., Gur, R. C., & Sachser, N. (2009). Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology, 115*(3), 257-263.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behav Brain Sci, 16*(1), 1-14. doi:10.1017/S0140525X00028636

Lopez, F. G., & Gormley, B. (2002). Stability and Change in Adult Attachment Style Over the First-Year College Transition: Relations to Self-Confidence, Coping, and Distress Patterns. *Journal of Counseling Psychology*, *49*(3), 355–364. https://doi.org/10.1037/0022-0167.49.3.355

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences, 205*(1161), 581-598.

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology, 8*(01), 4-15.

Grabe, E., & Low, E. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 515–546). Cambridge: Cambridge University Press.

Grannan-Rubenstein, G., Grannan-Rubenstein, W., & Thibodeau, P. (2014). Enculturation Effects of Musical Training on Pitch Discrimination.

Green, C. (2000). *Toddler Taming: A Parents' Guide to the First Four Years*. Vermillion.

Guedeney, A., Guedeney, N., Tereno, S., Dugravier, R., Greacen, T., Welniarz, B., Saias, T., & Tubach, F. (2011). Infant rhythms versus parental time: Promoting parent–infant synchrony. *Journal of Physiology - Paris*, *105*(4–6), 195–200. https://doi.org/10.1016/j.jphysparis.2011.07.005

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour, 42*(1), 1-14.

Gärdenfors, P. (2004). Cooperation and the Evolution of Symbolic Communication. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

Hammerschmidt, K., & Fischer, J. (2008). Constraints in primate vocal production. In D. K. Oller & U. Griebel (Eds.), *Evolution of communicative flexibility: complexity, creativity, and adaptability in human and animal communication* (pp. 93 – 119). Cambridge, MA: MIT Press.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*(11), 466-472. doi:10.1016/j.tics.2007.08.008

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science, 16*(1), 48-55.

Hardcastle, W. J., Laver, J., & Gibbon, F. E. (2010). *The handbook of phonetic sciences* (Vol. 1): John Wiley & Sons.

Hargreaves, D. J. (1986). Developmental Psychology and Music Education. *Psychology of Music, 14*(2), 83-96. doi:10.1177/0305735686142001

Harms, W. F. (2004). Primitive Content, Translation, and the Emergence of Meaning in Animal Communication. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

Harvey, P. H., & Clutton-Brock, T. H. (1985). Life history variation in primates. *Evolution*, 559-581.

Haydock, J., Koenig, W., & Stanback, M. (2001). Shared parentage and incest avoidance in the cooperatively breeding acorn woodpecker. *Molecular Ecology, 10*(6), 1515-1525.

Hayes, D. (2008). Tower a sign of the times. *Daily Variety, 301*(34), 1.

Hawkins, S. Personal communication, Santiago, 7-3-2017.

Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental Change in Auditory Preferences for Speech Stimuli in Japanese Infants. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1189–1200. https://doi.org/10.1044/1092-4388(2001/092)

Hazan, C., & Zeifman, D. (1994). Sex and the psychological tether. In *Advances in personal relationships: Vol. 5 Attachment processes in adulthood*. Jessica Kingsley Publishers.

Heffernan, M. E., & Fraley, R. C. (2015). How early experiences shape attraction, partner preferences, and attachment dynamics. In *Bases of adult attachment* (pp. 107-128): Springer.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology*, *6*.

Hennemann, W. W. (1984). Intrinsic rates of natural increase of altricial and precocial eutherian mammals: the potential price of precociality. *Oikos*, 363-368.

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5-to 7-month-olds in Vanuatu. *Developmental Science*, *22*(4), e12779.

Hepper, P. G. (1991). An examination of fetal learning before and after birth. *The Irish Journal of Psychology, 12*(2), 95-107.

Hesse, E. (1999). The Adult Attachment Interview: Historical and current perspectives. In *Handbook of attachment: Theory, research, and clinical applications* (pp. 395–433). Guilford Press.

Hewes, G. W., Andrew, R., Carini, L., Choe, H., Gardner, R. A., Kortlandt, A., . . . Pfeiffer, J. (1973). Primate communication and the gestural origin of language [and comments and reply]. *Current Anthropology, 14*(1/2), 5-24.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science (New York, N.Y.)*, *344*(6190), 1243091. https://doi.org/10.1126/science.1243091

Hinde, R. A. (1963). The nature of imprinting. In B. M. Foss (Ed.), *Determinants of infant behaviour* (Vol. 2, pp. 227-233). London: Methuen.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child development, 73*(2), 418-433.

Honing, H. (2011). *Musical cognition. A science of listening*. New Brunswick, NJ: Transaction Publishers.

Honing, H., Ladinig, O., Háden, G. P., & Winkler, I. (2009). Is Beat Induction Innate or Learned? *Annals of the New York Academy of Sciences, 11691*(1), 93-96. doi:10.1111/j.1749-6632.2009.04761.x

Honing, H., & Ploeger, A. (2012). Cognition and the evolution of music: Pitfalls and prospects. *Topics in cognitive science, 4*(4), 513-524.

Hoogland, J. L. (1983). Nepotism and alarm calling in the black-tailed prairie dog (Cynomys ludovicianus). *Animal Behaviour, 31*(2), 472-479. doi:https://doi.org/10.1016/S0003-3472(83)80068-2

House, J. (2006). Constructing a context with intonation. *Journal of Pragmatics, 38*(10), 1542-1558.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949–960.

Hrdy, S. (2001). Mothers and others. *Natural History, 110*(4), 50-63.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental psychology, 27*(2), 236.

Ilari, B. (2005). On musical parenting of young children: Musical beliefs and behaviors of mothers and infants. *Early Child Development and Care, 175*(7-8), 647-660.

Ilari, B. (2016). Music in the early years: Pathways into the social world. *Research Studies in Music Education*, *38*(1), 23–39. https://doi.org/10.1177/1321103X16642631

Insel, T. R., & Fernald, R. D. (2004). How the brain processes social information: Searching for the Social Brain. *Annual Review of Neuroscience, 27*(1), 697-722. doi:10.1146/annurev.neuro.27.070203.144148

Insel, T.R. & Young, L.J. (2001). The neurobiology of attachment. *Nature Reviews Neuroscience*, *2*(2), 129. https://doi.org/10.1038/35053579

Iriki, A., & Sakura, O. (2008). The neuroscience of primate intellectual evolution: natural selection and passive and intentional niche construction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1500), 2229-2241.

Iriki, A., & Taoka, M. (2012). Triadic (ecological, neural, cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367*(1585), 10-23.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An Introduction to Statistical Learning—with Applications in R*. London: Springer.

James, L. S., & Sakata, J. T. (2017). Learning Biases Underlie “Universals” in Avian Vocal Sequencing. *Current Biology, 27*(23), 3676-3682.

Janik, V. M., & Slater, P. J. (1997). Vocal learning in mammals. *Advances in the Study of Behaviour*, *26*, 59–100.

Jones, J. D., Cassidy, J., & Shaver, P. R. (2015). Adult attachment style and parenting. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and research: New directions and emerging themes*. London: The Guilford Press.

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and Rhythmic Patterns Affecting Infants Sensitivity to Musical Phrase Structure. *Journal of Experimental Psychology-Human Perception and Performance, 19*(3), 627-640.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). University Press.

Juslin, P. N., & Vastfjall, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behav Brain Sci*, *31*(5), 559–575; discussion 575-621. https://doi.org/10.1017/S0140525X08005293

Juslin, Patrik N, & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770.

Kagan, J. (1996). Three Pleasing Ideas. *American Psychologist*, *51*(9), 901–908. https://doi.org/10.1037/0003-066X.51.9.901

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., & Cooper, R. P. (1995). Dishabituation of visual attention by infant-versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior and Development, 18*(2), 209-223.

Kappeler, P. M., Silk, J. S., Burkart, J. M., & van Schaik, C. P. (2010). Primate behavior and human universals: Exploring the gap. In P. M. Kappeler & J. B. Silk (Eds.), *Mind the Gap* (pp. 3-15). Heidelberg: Springer.

Karantzas, G. C., & Simpson, J. A. (2015). Attachment and aged care. In *Attachment Theory and Research: New Directions and Emerging Themes* (pp. 319).

Katz, J. (1999). An episode of whining. In *How emotions work* (pp. 229–273). Chicago: University of Chicago Press.

Keefe, D. H., Bulen, J. C., Arehart, K. H., & Burns, E. M. (1993). Ear-canal impedance and reflection coefficient in human infants and adults. *The Journal of the Acoustical Society of America, 94*(5), 2617-2638. doi:10.1121/1.407347

Kempe, V. (2009). Child-directed speech prosody in adolescents: relationship to 2D: 4D, empathy, and attitudes towards children. *Personality and individual differences, 47*(6), 610-615.

Ko, E.-S. (2012). Nonlinear development of speaking rate in child-directed speech. *Lingua, 122*(8), 841-857.

Koelsch, S. (2010). From social contact to social cohesion: The 7 Cs. *Music & Medicine*, *5*(4), 204–219. https://doi.org/10.1177/1943862113508588

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., Huotilainen, M., & Fellman, V. (2010). Hemodynamic responses to speech and music in newborn infants. *Human Brain Mapping*, *31*(4), 595–603. https://doi.org/10.1002/hbm.20890

Kottak, C. P. (2017). *Window on Humanity: A Concise Introduction to General Anthropology* (8 ed.): McGraw-Hill Education.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). *Psychological Research on Joint Action* (Vol. 54, p. 101). Elsevier Science & Technology. https://doi.org/10.1016/B978-0-12-385527-5.00003-6

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & Von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters*, *6*(2), 156–159. https://doi.org/10.1098/rsbl.2009.0685

Kreutz, G., & Feldhaus, M. (2018). Does music help children grow up? Parental views from a longitudinal panel study. *Musicae Scientiae*, 102986491878258. doi:10.1177/1029864918782581

Lachmann, M., Szamado, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences, 98*(23), 13189-13194.

Lamont, A. (2008). Young children's musical worlds: musical engagement in 3.5-year-olds. *Journal of early childhood research, 6*(3), 247-261.

Langlois, J. H., Ritter, J. M., Casey, R. J., & Sawin, D. B. (1995). Infant attractiveness predicts maternal behaviors and attitudes. *Developmental Psychology, 31*(3), 464.

Laurent, H. K., & Ablow, J. C. (2012). The missing link: mothers' neural response to infant cry related to infant attachment behaviors. *Infant Behavior and Development, 35*(4), 761-772.

Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant behavior and development, 33*(4), 419-430.

Lee, C. S., Kitamura, C., Burnham, D., & Todd, N. P. M. (2014). On the rhythm of infant-versus adult-directed speech in Australian English. *The Journal of the Acoustical Society of America*, *136*(1), 357. https://doi.org/10.1121/1.4883479

LENA (2020). Building brains through early talk. Retrieved from the LENA website: https://www.lena.org/research/

Lerdahl, F., & Jackendoff, R. (1983). *A GenerativeTheory of Tonal Music* (1st ed.). MIT Press.

Levinson, S. C. (2006). *On the human" interaction engine".* Paper presented at the Wenner-Gren Foundation for Anthropological Research, Symposium 134.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philos Trans R Soc Lond B Biol Sci, 369*(1651), 20130302. doi:10.1098/rstb.2013.0302

Lewis, M. (1998). Altering Fate: Why the Past Does Not Predict the Future. *Psychological Inquiry*, *9*(2), 105–108. https://doi.org/10.1207/s15327965pli0902_6

Lewis, M., Feiring, C., & Rosenthal, S. (2000). Attachment over Time. *Child Development*, *71*(3), 707–720. https://doi.org/10.1111/1467-8624.00180

Lindsey, E. W., & Caldera, Y. M. (2015). Shared Affect and Dyadic Synchrony Among Secure and Insecure Parent–Toddler Dyads. *Infant and Child Development*, *24*(4), 394–413. https://doi.org/10.1002/icd.1893

Lipkind, D., & Tchernichovski, O. (2011). Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Supplement_3), 15572. https://doi.org/10.1073/pnas.1012941108

Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K. & Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, *498*(7452), 104. https://doi.org/10.1038/nature12173

Lobmaier, J. S., Sprengelmeyer, R., Wiffen, B., & Perrett, D. I. (2010). Female and male responses to cuteness, age and emotion in infant faces. *Evolution and Human Behavior, 31*(1), 16-21.

Longhi, E. (2009). Songese': Maternal structuring of musical interaction with infants. *Psychology of Music, 37*(2), 195-213.

Lorenz, K. (1970). Companions as factors in the bird's environment. *Studies in animal and human behavior, 1*, 101-258.

Luef, E., & Liebal, K. (2012). Infant-Directed Communication in Lowland Gorillas (Gorilla gorilla): Do Older Animals Scaffold Communicative Competence in Infants? *American journal of primatology, 74*(9), 841-852.

Luyten, P. (2015). Unholy questions about five central tenets of psychoanalysis that need to be empirically verified. *Psychoanalytic Inquiry*, *35*, 5–23.

Luyten, P., Fonagy, P., Lowyck, B., & Vermote, R. (2012). Assessment of mentalization. In *Handbook of mentalizing in mental health practice* (pp. 43–65). American Psychiatric Publishing.

Luyten, P., Mayes, L., Nijssens, L., & Fonagy, P. (2017). The parental reflective functioning questionnaire: Development and preliminary validation. *PLoS One*.

Luyten, P., Nijssens, L., Fonagy, P., & Mayes, L. (2017). Parental reflective functioning: Theory, research, and clinical applications. *The Psychoanalytic Study of the Child*.

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in Infancy, Childhood, and Adulthood: A Move to the Level of Representation. *Monographs of the Society for Research in Child Development*, *50*(1/2), 66. https://doi.org/10.2307/3333827

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth Strange Situation. *Attachment in the preschool years: Theory, research, and intervention, 1*, 121-160.

Maizels, N. (1985). Self-Envy, the Womb and the Nature of Goodness—A Reappraisal of the Death Instinct. *International Journal of Psycho-Analysis*, *66*, 185–192.

Malloch, S. (2000). Mothers and infants and communicative musicality. *Musicae Scientiae: Journal of the European Society for the Cognitive Sciences of Music, Special Issue*(1999–2000), 29–57.

Malloch, S., & Trevarthen, C. (2008). *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford: Oxford University Press.

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 1-10). New York: Oxford University Press.

Marcello, A. (2006). Contemplating the evolution of attachment and cognition in a collaborative learning environment.

Marsh, K., Young, S., & McPherson, G. (2016). Musical play. In *The child as musician: A handbook of musical development* (2nd ed., pp. 462–484). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198744443.003.0025

Masataka, N. (1999). Preference for infant-directed singing in 2-day-old hearing infants of deaf parents. *Developmental psychology, 35*(4), 1001.

Matthews, G. V. T. (1969). Navigation in Animals. *J. Navigation, 22*(1), 118-126. doi:10.1017/S0373463300041060

McConnell-Ginet, S. (1978). Intonation in a man's world. *Signs: Journal of Women in Culture and Society, 3*(3), 541-559.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception, 7*(5), 491-497.

Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-month-old infants, melodies are social. *Psychological Science*, *27*(4), 486–501.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings,

M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science (New York, N.Y.)*, *366*(6468). https://doi.org/10.1126/science.aax0868

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*.

Mertens, P. (2004). *The prosogram: Semi-automatic transcription of prosody based on a tonal perception model.* Paper presented at the Speech Prosody 2004, International Conference.

Meunier, H. (2016). Do monkeys have a theory of mind? How to answer the question? *Neuroscience & Biobehavioral Reviews*.

Mikulincer, M., & Shaver, P. R. (2007). *Attachment in Adulthood. Structure, Dynamics, and Change*. The Guilford Press.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child development, 78*(2), 622-646.

Mitchell, R. L., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience & Biobehavioral Reviews, 37*(3), 471-479.

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind, and body*. Cambridge (MA): Harvard University Press.

Montague, P. R., & Lohrenz, T. (2007). To Detect and Correct: Norm Violations and Their Enforcement. *Neuron*, *56*(1), 14–18. https://doi.org/10.1016/j.neuron.2007.09.020

Montuori, L. M., & Honey, R. C. (2015). Representation in development: From a model system to some general processes. *Neuroscience & Biobehavioral Reviews, 50*, 143-149. doi:https://doi.org/10.1016/j.neubiorev.2014.03.005

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111*(981), 855-869.

Mote, J. (2011). The effects of tempo and familiarity on children's affective interpretation of music. *Emotion, 11*(3), 618-622.

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455–464.

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music, 31*(1), 75-92.

Neumann, I. D. (2008). Brain Oxytocin: A Key Regulator of Emotional and Social Behaviours in Both Females and Males. *Journal of Neuroendocrinology*, *20*(6), 858–865. https://doi.org/10.1111/j.1365-2826.2008.01726.x

Newman, J. D. (2003). Vocal communication and the triune brain. *Physiology & behavior, 79*(3), 495-502.

Newman, R. S., & Hussain, I. (2006). Changes in Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy*, *10*(1), 61–76. https://doi.org/10.1207/s15327078in1001_4

Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition*, *52*(3), 382–389. https://doi.org/10.1016/S0278-2626(03)00182-9

Nolan, F., & Asu, E. L. (2009). The Pairwise Variability Index and Coexisting Rhythms in Language. *Phonetica*, *66*(1–2), 64–77. https://doi.org/10.1159/000208931

Nuttall, P. (2008). For Those About to Rock: a New Understanding of Adolescent Music Consumption. *Advances in Consumer Research, 35*, 1.

Ohala, J. J. (1980). The acoustic origin of the smile. *The Journal of the Acoustical Society of America, 68*(S1), S33-S33.

Oller, D. K. (2004). Underpinnings for a Theory of Communicative Evolution. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

O'Neill, D. K. (2007). The Language Use Inventory: A parent-report measure of pragmatic language development for 18-47-month-old children. *Journal of Speech, Language and Hearing Research*, *50*, 214–228.

Ostwald, P. (1972). The sounds of infancy. *Developmental Medicine & Child Neurology, 14*(3), 350-361.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences, 106*(51), 22026-22031. doi:10.1073/pnas.0908118106

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE, 4*(11), e7808. doi:10.1371/journal.pone.0007808

Overall, N. C., Fletcher, G. J. O., Simpson, J. A., & Fillo, J. (2015). Attachment insecurity, biased perceptions of romantic partners' negative emotions, and hostile relationship behavior. *Journal of Personality and Social Psychology, 108*(5), 730-749.

Owings, D. H., & Morton, E. S. (1997). The role of information in communication: A management/assessment approach. In D. H. Owings, M. D. Beecher, & N. S. Thompson (Eds.), *Perspectives in ethology, XII: Communication* (pp. 359–390): Plenum Press.

Owings, D. H., & Zeifman, D. M. (2004). Human Infant Crying as an Animal Communication System: Insights from an Assessment/Management Approach. In D.

K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews, 10*(2), 58-71.

Palmer, C., & Hutchins, S. (2006). What is musical prosody? In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 46, pp. 245–278). Elsevier Academic Press.

Panksepp, J. (1995). The emotional sources of" chills" induced by music. *Music Perception: An Interdisciplinary Journal, 13*(2), 171-207.

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: do animals have affective lives? *Neuroscience & Biobehavioral Reviews, 35*(9), 1791-1804.

Papousek, M. (1996). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliege & J. Sloboda (Eds.), *Musical beginnings: origins and development of musical competence* (pp. 37-55). Oxford: Oxford University Press.

Papousek, M., & Papousek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language, 9*(6), 137-157. doi:10.1177/014272378900900603

Pasiali, V. (2014). Music therapy and attachment relationships across the life span. *Nordic Journal of Music Therapy*, *23*(3), 202–223. https://doi.org/10.1080/08098131.2013.829863

Patel, A. D. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing Prosodic and Musical Patterns: A Neuropsychological Investigation. Brain and Language, 61(1), 123–144. https://doi.org/10.1006/brln.1997.1862

Pearce, C. (2009). *A short introduction to attachment and attachment disorder*. London: Jessica Kingsley Publishers.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107-120.

Pentland, A. (2010). To Signal Is Human: Real-time data mining unmasks the power of imitation, kith and charisma in our face-to-face social networks. *American scientist, 98*(3), 204-211.

Pentland, A., & Heibeck, T. (2008). *Honest signals: how they shape our world*. London: MIT Press.

Penton-Voak, I. S., Cahill, S., Pound, N., Kempe, V., Schaeffler, S., & Schaeffler, F. (2007). Male facial attractiveness, perceived personality, and child-directed speech. *Evolution and Human Behavior, 28*(4), 253-259.

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Front Hum Neurosci, 6*, 26. doi:10.3389/fnhum.2012.00026

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529): Penguin UK.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and brain sciences, 13*(04), 707-727.

Pinquart, M., Feußner, C., & Ahnert, L. (2013). Meta-analytic evidence for stability in attachments from infancy to early adulthood. *Attachment & Human Development*, *15*(2), 189–218. https://doi.org/10.1080/14616734.2013.746257

Porta, M. (2014). Qualitative Research. In (6 ed.): Oxford University Press.

Povinelli, D. (1996). Chimpanzee theory of mind?: the long road to strong inference. In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind* (pp. 293-329). Cambridge: Cambridge University Press.

Querleu, D., Renard, X., & Crépin, G. (1981). [Auditory perception and fetal reaction to react to sound stimulation (author's transl)]. *Journal de gynecologie, obstetrique et biologie de la reproduction, 10*(4), 307-314.

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America, 113*(6), 3390-3402. doi:10.1121/1.1568942

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour, 78*(2), 233-240.

Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences, 24*(02), 309-324.

Reyna, B. A., & Pickler, R. H. (2009). Mother-Infant Synchrony. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, *38*(4), 470–477. https://doi.org/10.1111/j.1552-6909.2009.01044.x

Rholes, W., & Simpson, J. A. (2004). *Adult attachment: Theory, research, and clinical implications*: Guilford Publications.

Rijt-Plooij, H. H., & Plooij, F. X. (1993). Distinct Periods of Mother-Infant Conflict in Normal Development: Sources of Progress and Germs of Pathology. *Journal of child psychology and psychiatry, 34*(2), 229-245. Bateson

Roisman, G. I., Collins, W. A., Sroufe, L. A., & Egeland, B. (2005). Predictors of young adults' representations of and behavior in their current romantic relationship: Prospective tests of the prototype hypothesis. *Attachment & Human Development*, *7*(2), 105–121. https://doi.org/10.1080/14616730500134928

Roisman, G. I., Holland, A., Fortuna, K., Fraley, R. C., Clausell, E., & Clarke, A. (2007). The Adult Attachment Interview and Self-Reports of Attachment Style: An Empirical Rapprochement. *Journal of Personality and Social Psychology*, *92*(4), 678–697. https://doi.org/10.1037/0022-3514.92.4.678

Rosenberg, K., & Trevathan, W. (1995). Bipedalism and human birth: The obstetrical dilemma revisited. *Evolutionary Anthropology: Issues, News, and Reviews, 4*(5), 161-168.

Ruff, C. B., & Walker, A. (1993). Body size and body shape. *The Nariokotome Homo erectus skeleton*, 234-265.

Russell, B. (1940). *An Inquiry into Meaning and Truth*. Allen & Unwin.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, 107*(6), 2408-2412.

Sayigh, L. S., Tyack, P. L., Wells, R. S., & Scott, M. D. (1990). Signature whistles of free-ranging bottlenose dolphins Tursiops truncatus: stability and mother-offspring comparisons. *Behavioral Ecology and Sociobiology, 26*(4), 247-260.

Schaffer, H. R., & Emerson, P. E. (1964). The Development of Social Attachments in Infancy. *Monographs of the Society for Research in Child Development, 29*(3), 1-77. doi:10.2307/1165727

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology, 32*(1), 76-92.

Schoenherr, D., Strauss, B., Paulick, J., Deisenhofer, A.-K., Schwartz, B., Rubel, J. A., Boyle, K., Lutz, W., Stangier, U., & Altmann, U. (2019). Movement Synchrony and Attachment Related Anxiety and Avoidance in Social Anxiety Disorder. *Journal of Psychotherapy Integration*. https://doi.org/10.1037/int0000187

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A Domain-General Theory of the Development of Perceptual Discrimination. *Current Directions in Psychological Science*, *16*(4), 197–201. https://doi.org/10.1111/j.1467-8721.2007.00503.x

Scott-Phillips, T. C. (2008). Defining biological communication. *Journal of evolutionary biology, 21*(2), 387-395.

Seddon, N., Tobias, J. A., & Alvarez, A. (2002). Vocal communication in the pale-winged trumpeter (Psophia leucoptera): repertoire, context and functional reference. *Behaviour, 139*(10), 1331-1359.

Seeger, A. (2004). *Why Suyá Sing: A Musical Anthropology of an Amazonian People*. Champaign, IL: University of Illinois Press.

Seifritz, E., Esposito, F., Neuhoff, J. G., Lüthi, A., Mustovic, H., Dammann, G., . . . Tedeschi, G. (2003). Differential sex-independent amygdala response to infant crying and laughing in parents versus nonparents. *Biological psychiatry, 54*(12), 1367-1375.

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, *108*(2), 303–319. https://doi.org/10.1016/j.cognition.2008.02.009

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual review of psychology, 54*(1), 145-173.

Seyfarth, R. M., Cheney, D. L., & Bergman, T. J. (2005). Primate social cognition and the origins of language. *Trends in Cognitive Sciences, 9*(6), 264-266. doi:https://doi.org/10.1016/j.tics.2005.04.001

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science, 210*(4471), 801-803.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour, 28*(4), 1070-1094. doi:https://doi.org/10.1016/S0003-3472(80)80097-2

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, Il.: University of Illinois Press.

Shannon, K. (2006). Infant behavioral responses to infant-directed singing and other maternal interactions. *Infant Behavior and Development, 29*(3), 456-470.

Sharp, S. P., & Hatchwell, B. J. (2006). Development of family specific contact calls in the Long-tailed Tit Aegithalos caudatus. *Ibis, 148*(4), 649-656.

Shea, B. T. (1987). Reproductive strategies, body size, and encephalization in primate evolution. *International journal of primatology, 8*(2), 139-156.

Silk, J. B., & Boyd, R. (2010). From grooming to giving blood: the origins of human altruism. In P. M. Kappeler & J. B. Silk (Eds.), *Mind the Gap* (pp. 223-244). Heidelberg: Springer.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3*(3), 365-394.

Sinha, C. (2004). The Evolution of Language: From Signals to Symbols to System. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach*. Cambridge, MA: MIT Press.

Sininger, Y. S., Abdala, C., & Cone-Wesson, B. (1997). Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. *Hearing research, 104*(1), 27-38.

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: A preliminary study. *Attachment & Human Development*, *7*(3), 283–298. https://doi.org/10.1080/14616730500245880

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Animal Behaviour, 81*(5), 919-924.

Small, C. (1998). *Musicking The Meanings of Performing and Listening / Christopher Small.* Hanover : University Press of New England.

Smith, J. M., & Harper, D. (2003). *Animal signals*: Oxford University Press.

Smith, N. A., & Trainor, L. J. (2008). Infant-Directed Speech Is Modulated by Infant Feedback. *Infancy, 13*(4), 410-420. doi:10.1080/15250000802188719

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501-532. doi:http://dx.doi.org/10.1016/j.dr.2007.06.002

Sokol, R. I., Webster, K. L., Thompson, N. S., & Stevens, D. A. (2005). Whining as mother-directed speech. *Infant and Child Development, 14*(5), 478-490.

Sole, M. (2016). Crib song: Insights into functions of toddlers' private spontaneous singing. *Psychology of Music*, *45*(2), 172–192. https://doi.org/10.1177/0305735616650746

Soltis, J. (2004). The signal functions of early infant crying. *Behavioral and Brain Sciences, 27*(04), 443-458.

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development, 10*(2), 133-142.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 1). Oxford: Blackwell.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Sperber, D., & Wilson, D. (2010). Précis of Relevance: Communication and Cognition. *Behavioral and Brain Sciences, 10*(4), 697-710.

Sprengelmeyer, R., Perrett, D. I., Fagan, E., Cornwell, R., Lobmaier, J., Sprengelmeyer, A., . . . Crow, S. (2009). The cutest little baby face: A hormonal link to sensitivity to cuteness in infant faces. *Psychological Science, 20*(2), 149-154.

Sroufe, A., Egeland, B., Carlson, E., Collins, A., & Thompson, R. (2006). The development of the person: The Minnesota study of risk and adaptation from birth to adulthood. *Human Development*, *49*(2), 138–142.

Sroufe, L., Egeland, B., & Kreutzer, T. (1990). The Fate of Early Experience Following Developmental Change: Longitudinal Approaches to Individual Adaptation in Childhood. *Child Development*, *61*(5), 1363–1373.

Starck, J. M. (1998). *Avian growth and development: evolution within the altricial-precocial spectrum*: Oxford University Press on Demand.

Street, A., Young, S., Tafuri, J., & Ilari, B. (2003). *Mothers' attitudes to singing to their infants.* Paper presented at the Proceedings of the 5th ESCOM conference, Hanover University of Music and Drama, Germany.

Stuart, P. S., Andrew, M., Matthew, J. W., & Ben, J. H. (2005). Learned kin recognition cues in a social bird. *Nature, 434*(7037), 1127. doi:10.1038/nature03522

Sun, H., & Jensen, R. (1994). Body segment growth during infancy. *Journal of Biomechanics, 27*(3), 265-275.

Sundberg, J. (1987). *The Science of the Singing Voice*. Delkalb, Illinois.: Northern Illinois University Press.

Swain, J. E., Mayes, L. C., & Leckman, J. F. (2004). The development of parent-infant attachment through dynamic and interactive signaling loops of care and cry. *Behavioral and Brain Sciences, 27*(4), 472-473.

Számadó, S. (2012). The rise and fall of handicap principle: a commentary on the “Modelling and the fall and rise of the handicap principle”. *Biology & Philosophy, 27*(2), 279-286.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & psychophysics, 27*(1), 24-27.

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *The Journal of the Acoustical Society of America, 96*(4), 2101-2107.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)*, *4*(1), 46.

Tierney, A. T., Russo, F. A., & Patel, A. D. (2008). Empirical comparisons of pitch patterns in music, speech, and birdsong. *Journal of the Acoustical Society of America, 123*(5), 37-21.

Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language Sciences, 10*(1), 69-88.

Tomasello, M., & Camaioni, L. (1997). A comparison of the gestural communication of apes and human infants. *Human Development, 40*(1), 7-24.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and brain sciences, 16*(03), 495-511.

Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development, 19*(1), 83-92.

Trainor, L. J. (2006). Innateness, Learning, and the Difficulty of Determining Whether Music is an Evolutionary Adaptation. A Commentary on Justus & Hutsler (2005) and McDermott & Hauser (2005). *Music Perception: An Interdisciplinary Journal, 24*(1), 105-110. doi:10.1525/mp.2006.24.1.105

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

Trask, R. L. (2007). *Language and linguistics: The key concepts*: Taylor & Francis.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature neuroscience, 6*(7), 669-673.

Trehub, S. E. (2010). Musicality in the eye or ear of the beholder. *Psychology of Music, 38*(4), 499-502.

Trehub, S. E. (2016). Infant musicality. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford handbook of music psychology* (2nd ed., pp. 387-397). Oxford: Oxford University Press.

Trehub, S. E., & Degé, F. (2016). Reflections on infants as musical connoisseurs. In G. E. McPherson (Ed.), *The Child as Musician A handbook of musical development* (2 ed., pp. 31-51). Oxford: Oxford University press.

Trehub, S. E., & Nakata, T. (2002). Emotion and music in infancy. *Musicae Scientiae, 5*(1 suppl), 37-61.

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. *Developmental psychology, 33*(3), 500.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant behavior and development, 16*(3), 285-295.

Trevarthen, C. (1993). The self born in intersubjectivity: The psychology of an infant communicating. In U. Neisser (Ed.), *The perceived self: Ecological and interpersonal sources of self-knowledge*. Cambridge: Cambridge University Press.

Trivers, R. L. (1974). Parent-offspring conflict. *American zoologist, 14*(1), 249-264.

Tuckett, A. G. (2005). Applying thematic analysis theory to practice: A researcher's experience. *Contemporary Nurse*, *19*, 75–87.

Tunçgenç, B., Cohen, E., & Fawcett, C. (2015). Rock With Me: The Role of Movement Synchrony in Infants' Social and Nonsocial Choices. *Child Development*, *86*(3), 976–984. https://doi.org/10.1111/cdev.12354

Turino, T. (2008). *The Politics of Participation*. Chicago: University of Chicago Press.

Ullstadius, E. (2014). Conflict development in infancy. *Learning, Culture and Social Interaction, 3*(3), 210-216. doi:https://doi.org/10.1016/j.lcsi.2014.02.014

Van der Horst, F. (2011). John Bowlby–From Psychoanalysis to Ethology. Unraveling the Roots of Attachment Theory. Chichester (Wiley) 2011.

Van Puyvelde, M., & Franco, F. (2015). The interaction of music and language in the ontogenesis of human communication: a multimodal parent-infant co-regulation system.

Van Puyvelde, M., Vanfleteren, P., Loots, G., Deschuyffeleer, S., Vinck, B., Jacquet, W., & Verhelst, W. (2010). Tonal synchrony in mother–infant interaction based on harmonic and pentatonic series. *Infant Behavior and Development, 33*(4), 387-400. doi:http://dx.doi.org/10.1016/j.infbeh.2010.04.003

van Schaik, C. P., & Burkart, J. M. (2010). Mind the Gap: Cooperative Breeding and the Evolution of Our Unique Features. In P. M. Kappeler & J. B. Silk (Eds.), *Mind the Gap*. Heidelberg: Springer.

Vicaria, I., & Dickens, L. (2016). Meta-Analyses of the Intra- and Interpersonal Outcomes of Interpersonal Coordination. *Journal of Nonverbal Behavior*, *40*(4), 335–361. https://doi.org/10.1007/s10919-016-0238-8

Vicedo, M. (2009). The father of ethology and the foster mother of ducks: Konrad Lorenz as expert on motherhood. *Isis, 100*(2), 263-291.

Vitouch, O., & Ladinig, O. (2009). Music and evolution. *Musicae Scientiae, 13*(2), 7-11.

Vosoughi, S., & Roy, D. K. (2012). A longitudinal study of prosodic exaggeration in child-directed speech.

Vuoskoski, J. K., Clarke, E. F., & DeNora, T. (2016). Music listening evokes implicit affiliation. *Psychology of Music*, 0305735616680289.

Wallman, J. (1992). *Aping language*. New York: Cambridge University Press.

Walters, J. R. (1987). Kin recognition in non-human primates. In D. J. C. Fletcher & C. D. Michener (Eds.), *Kin recognition in animals* (pp. 359-393). New York: Wiley.

Wanker, R., Apcin, J., Jennerjahn, B., & Waibel, B. (1998). Discrimination of different social companions in spectacled parrotlets (Forpus conspicillatus): evidence for individual vocal recognition. *Behavioral Ecology and Sociobiology, 43*(3), 197-202.

Warren-Leubecker, A., & Bohannon III, J. N. (1984). Intonation patterns in child-directed speech: Mother-father differences. *Child Development*, 1379-1385.

Waters, E., Merrick, S., Treboux, D., Crowell, J., & Albersheim, L. (2000). Attachment Security in Infancy and Early Adulthood: A Twenty-Year Longitudinal Study. *Child Development*, *71*(3), 684–689. https://doi.org/10.1111/1467-8624.00176

Whitehead, R. (1985). The human weaning process. *Pediatrics, 75*(1), 189-193.

Whiten, A., & Byrne, R. W. (1988). The Machiavellian Intelligence hypothesis. In R. W. Byrne & A. Whiten (Eds.), *Machiavellian Intelligence*. Oxford: Oxford University Press.

Williams, K. E., Barrett, M. S., Welch, G. F., Abad, V., & Broughton, M. (2015). Associations between early shared music activities in the home and later child outcomes: Findings from the Longitudinal Study of Australian Children. *Early Childhood Research Quarterly*, *31*, 113–124. https://doi.org/10.1016/j.ecresq.2015.01.004

Wilson, M., & Cook, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic bulletin & review*, 1-13.

Winkler, I., Ha´den, G., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences, 106*, 2468–2471.

Wu, R., Tummeltshammer, K., Gliga, T., & Kirkham, N. (2014). Ostensive signals support learning from novel attention cues during infancy. *Frontiers in Psychology*, *5*, 251. https://doi.org/10.3389/fpsyg.2014.00251

Yee, C. (2015). *Synchrony and Attachment*. Arizona State University.

Young, E. S., Simpson, J. A., Griskevicius, V., Huelsnitz, C. O., & Fleck, C. (2017). Childhood attachment and adult personality: A life history perspective. *Self and Identity*, 1-17.

Young, S. (2005). Changing Tune: Reconceptualizing Music with Under Three Year Olds. *International Journal of Early Years Education, 13*(3), 289-303. doi:10.1080/09669760500295987

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *Journal of theoretical biology, 67*(3), 603-605. doi:10.1016/0022-5193(77)90061-3

Zebrowitz, L. A., Brownlow, S., & Olson, K. (1992). Baby talk to the babyfaced. *Journal of Nonverbal Behavior, 16*(3), 143-158.

Zeifman, D. M. (2001). An ethological analysis of human infant crying: Answering Tinbergen's four questions. *Developmental psychobiology, 39*(4), 265-285.

Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences, 107*(13), 5768-5773.

Wolff, M. S., & Ijzendoorn, M. H. (1997). Sensitivity and Attachment: A Meta-Analysis on Parental Antecedents of Infant Attachment. *Child Development*, *68*(4), 571–591. https://doi.org/10.1111/j.1467-8624.1997.tb04218.x

Zuberbühler, K. (2008). Audience effects. *Current Biology, 18*(5), R189-R190.

APPENDICES

Appendix A

Set of questions prepared in order to assess the first study's participants' perceptions, opinions, beliefs, and attitudes towards the IDReg and related topics.

Preparation

- Set Canon Legria & Iphone
  - Both slots empty
  - Recording in low quality
- Timer
- Attendee list and register
- Labels for names
- All forms in table (ethical consent, participant information, contact sheet, incentive log)

Setting

- Greet
- Consent form
- Participant information sheet

1. Briefing:

- Welcome and thanks
- Timings

Time	Topic
15'	Briefing
10'	Ice-breaker
15'	Debriefing

- Roles - expectations and responsibilities for participants and facilitators ('rules of the game')
  - I'll be moderating
  - I'll ask general questions about parenting
  - essentially a group conversation, each one share impressions, experiences. Don't have to be 100% sure
  - respect each other, no judging
  - might moderate someone in order to let others talk, don't take offence
- How information from focus group will be analyzed and used
- Opportunity for participants to ask any questions and talk in confidence
- Start recording

2.Ice breaking

- Introduce researcher
- Let members introduce themselves:
  - *Who are you? Occupation? Hobby?*
  - *How many children do you have? How old?*

3. Conversation

- *Do you know what IDSp (motherese, babytalk, etc.) is?*

(If nobody answers, introduce the general topic

What IDReg is (isn't)

(It seems like parents across cultures do not speak with their children using the same vocal range or register throughout the day. They sometimes switch between registers rather seamlessly, phenomenon called code-switching)

- *Why do you use it?*
- *How does it feel?*
- *Do you think it has an impact in the interaction?*

- **Has the introduction of language changed your relationship with your child?**
- **If so, how?**
- **Has your child's ability to talk had implications in the way you address him?**

- *When do you still use it?*
- *Do your use of IDSp depends on the instance?*
- In which instances do you used it/not use it?
    - telling her how cute he/she is
    - calling for a bath
    - playing
    - when responding to something 'through the babies eyes'
    - case of illness or injury
    - bedtime
    - waking up)

If they do remember a qualitative change on the use of IDSp
Do you remember any request from the child to maintain IDSp?
Do you remember any emotional reaction to the use of 'drier' ADSp?
Do you remember any change in the child's sensitivity to IDSp? To music?

Do you sometimes avoid it?
In which circumstances would you avoid it?

Does your use of IDSp change when other people are present?
Have you noticed any change in the nature or IDSp or CS rate when another child is present?
Do you feel Inhibition?
Are sometimes (older) siblings jealous of the use of IDSp to a child?
Have you noticed jealousy in surrounding siblings when addressing one of your children in IDSp?

4. Debriefing:
- What happens next: contact sheet
- Offer opportunity for them to ask any final questions
- Thank them for their time
- Arrange incentives

Appendix B.

Second study's Initial questionnaire, concerning basic demographics and participating parents' own holistic appraisal of their participating child's development.

Participant information sheet

Name _________________________ Surname ________________________________

Age ____

Email (if not previously provided) ____

Latest Qualification ____

Occupation and number of hours a week _______________________________________

____

Average daily hours spent with target child _____________________________________

Mother tongue ___________________

Language(s) spoken to target child and average proportion (e.g. Italian 30%)

____

Number of Children ____

Age of Child(ren) (in years and months) __________________________________

Screening questions

1. How much do you agree with the following claims?
(-3 = I do not agree at all, 3= I completely agree)

1.1. My child understands everything I say to her/him ___

1.2. My child understands sentences (not just single words, but combinations of them) ___

1.3. Sometimes, my child understands me, but pretends like she/he does not ___

1.4. Sometimes my child does not cooperate due to the fact that he/she does not understand what I'm asking him/her ___

1.5. My child has learned to walk ___

2. What does her/his locomotion look like? (e.g. crawling, etc.)

____
____

3. How autonomous would you say your target child is?
(-3 = Very dependent, 3 = Very autonomous)

____

Appendix C.

MacArthur-Bates Communicative Development Inventories.

MacArthur-Bates
Communicative Development Inventories

Child's name: ______________________ Gender: ______________________

Birthdate: ______________________ Today's date: ______________________

MacArthur-Bates CDI
Words and Sentences

Distributed by Paul H. Brookes Publishing Co.
1-800-638-3775; 410-337-9580
www.brookespublishing.com

Larry Fenson, Ph.D.
Virginia A. Marchman, Ph.D.
Donna J. Thal, Ph.D.
Philip S. Dale, Ph.D.
J. Steven Reznick, Ph.D.
Elizabeth Bates, Ph.D.

Part I: Words Children Use

A. Vocabulary Checklist

Children understand many more words than they say. We are particularly interested in the words your child SAYS. Please go through the list and mark the words you have heard your child use. If your child uses a different pronunciation of a word (for example, "raffe" instead of "giraffe" or "sketti" for "spaghetti"), mark the word anyway. Remember that this is a "catalogue" of all the words that are used by many different children. Don't worry if your child knows only a few of these right now.

1. *Sound Effects and Animal Sounds* (12)

baa baa	○	grrr	○	ouch	○	vroom	○
choo choo	○	meow	○	quack quack	○	woof woof	○
cockadoodledoo	○	moo	○	uh oh	○	yum yum	○

2. *Animals (Real or Toy)* (43)

alligator	○	cow	○	horse	○	puppy	○
animal	○	deer	○	kitty	○	rooster	○
ant	○	dog	○	lamb	○	sheep	○
bear	○	donkey	○	lion	○	squirrel	○
bee	○	duck	○	monkey	○	teddybear	○
bird	○	elephant	○	moose	○	tiger	○
bug	○	fish	○	mouse	○	turkey	○
bunny	○	frog	○	owl	○	turtle	○
butterfly	○	giraffe	○	penguin	○	wolf	○
cat	○	goose	○	pig	○	zebra	○
chicken	○	hen	○	pony	○		

3. *Vehicles (Real or Toy)* (14)

airplane	○	car	○	sled	○	tricycle	○
bicycle	○	firetruck	○	stroller	○	truck	○
boat	○	helicopter	○	tractor	○		
bus	○	motorcycle	○	train	○		

4. *Toys* (18)

ball	○	bubbles	○	glue	○	puzzle	○
balloon	○	chalk	○	pen	○	story	○
bat	○	crayon	○	pencil	○	toy	○
block	○	doll	○	play dough	○		
book	○	game	○	present	○		

5. *Food and Drink* (68)

apple ○	corn ○	lollipop ○	pretzel ○
applesauce ○	cracker ○	meat ○	pudding ○
banana ○	donut ○	melon ○	pumpkin ○
beans ○	drink ○	milk ○	raisin ○
bread ○	egg ○	muffin ○	salt ○
butter ○	fish ○	noodles ○	sandwich ○
cake ○	food ○	nuts ○	sauce ○
candy ○	french fries ○	orange ○	soda/pop ○
carrots ○	grapes ○	pancake ○	soup ○
cereal ○	green beans ○	peanut butter ○	spaghetti ○
cheerios ○	gum ○	peas ○	strawberry ○
cheese ○	hamburger ○	pickle ○	toast ○
chicken ○	ice ○	pizza ○	tuna ○
chocolate ○	ice cream ○	popcorn ○	vanilla ○
coffee ○	jello ○	popsicle ○	vitamins ○
coke ○	jelly ○	potato ○	water ○
cookie ○	juice ○	potato chip ○	yogurt ○

6. *Clothing* (28)

beads ○	dress ○	pajamas ○	sneaker ○
belt ○	gloves ○	pants ○	snowsuit ○
bib ○	hat ○	scarf ○	sock ○
boots ○	jacket ○	shirt ○	sweater ○
button ○	jeans ○	shoe ○	tights ○
coat ○	mittens ○	shorts ○	underpants ○
diaper ○	necklace ○	slipper ○	zipper ○

7. *Body Parts* (27)

ankle ○	eye ○	knee ○	shoulder ○
arm ○	face ○	leg ○	tooth ○
belly button ○	feet ○	lips ○	toe ○
buttocks/bottom* ○	finger ○	mouth ○	tongue ○
cheek ○	hair ○	nose ○	tummy ○
chin ○	hand ○	owie/boo boo ○	vagina* ○
ear ○	head ○	penis* ○	

*or word used in your family

Small Household Items (50)

basket	dish	mop	spoon
blanket	fork	nail	tape
bottle	garbage	napkin	telephone
box	glass	paper	tissue/kleenex
bowl	glasses	penny	toothbrush
broom	hammer	picture	towel
brush	jar	pillow	trash
bucket	keys	plant	tray
camera	knife	plate	vacuum
can	lamp	purse	walker
clock	light	radio	watch
comb	medicine	scissors	
cup	money	soap	

Furniture and Rooms (33)

basement	crib	play pen	stairs
bathroom	door	porch	stove
bathtub	drawer	potty	table
bed	dryer	refrigerator	TV
bedroom	garage	rocking chair	washing machine
bench	high chair	room	window
chair	kitchen	shower	
closet	living room	sink	
couch	oven	sofa	

Outside Things (31)

backyard	lawn mower	sidewalk	stone
cloud	moon	sky	street
flag	pool	slide	sun
flower	rain	snow	swing
garden	rock	snowman	tree
grass	roof	sprinkler	water
hose	sandbox	star	wind
ladder	shovel	stick	

11. *Places to Go* (22)

beach ◯	farm ◯	park ◯	woods ◯
camping ◯	gas station ◯	party ◯	work ◯
church* ◯	home ◯	picnic ◯	yard ◯
circus ◯	house ◯	playground ◯	zoo ◯
country ◯	movie ◯	school ◯	
downtown ◯	outside ◯	store ◯	

*or word used in your family

12. *People* (29)

aunt ◯	cowboy ◯	lady ◯	pet's name ◯
baby ◯	daddy* ◯	mailman ◯	police ◯
babysitter ◯	doctor ◯	man ◯	sister ◯
babysitter's name ◯	fireman ◯	mommy* ◯	teacher ◯
boy ◯	friend ◯	nurse ◯	uncle ◯
brother ◯	girl ◯	child's own name ◯	
child ◯	grandma* ◯	people ◯	
clown ◯	grandpa* ◯	person ◯	

*or word used in your family

13. *Games and Routines* (25)

bath ◯	go potty ◯	patty cake ◯	thank you ◯
breakfast ◯	hi ◯	peekaboo ◯	this little piggy ◯
bye ◯	hello ◯	please ◯	turn around ◯
call (on phone) ◯	lunch ◯	shh/shush/hush ◯	yes ◯
dinner ◯	nap ◯	shopping ◯	
give me five! ◯	night night ◯	snack ◯	
gonna get you! ◯	no ◯	so big! ◯	

14. *Action Words* (103)

bite ◯	catch ◯	cry ◯	dump ◯
blow ◯	chase ◯	cut ◯	eat ◯
break ◯	clap ◯	dance ◯	fall ◯
bring ◯	clean ◯	draw ◯	feed ◯
build ◯	climb ◯	drink ◯	find ◯
bump ◯	close ◯	drive ◯	finish ◯
buy ◯	cook ◯	drop ◯	fit ◯
carry ◯	cover ◯	dry ◯	fix ◯

(continued)

get	◯	listen	◯	see	◯	take	◯
give	◯	look	◯	shake	◯	talk	◯
go	◯	love	◯	share	◯	taste	◯
hate	◯	make	◯	show	◯	tear	◯
have	◯	open	◯	sing	◯	think	◯
hear	◯	paint	◯	sit	◯	throw	◯
help	◯	pick	◯	skate	◯	tickle	◯
hide	◯	play	◯	sleep	◯	touch	◯
hit	◯	pour	◯	slide	◯	wait	◯
hold	◯	pretend	◯	smile	◯	wake	◯
hug	◯	pull	◯	spill	◯	walk	◯
hurry	◯	push	◯	splash	◯	wash	◯
jump	◯	put	◯	stand	◯	watch	◯
kick	◯	read	◯	stay	◯	wipe	◯
kiss	◯	ride	◯	stop	◯	wish	◯
knock	◯	rip	◯	sweep	◯	work	◯
lick	◯	run	◯	swim	◯	write	◯
like	◯	say	◯	swing	◯		

15. *Descriptive Words* (63)

allgone	◯	dry	◯	last	◯	scared	◯
asleep	◯	empty	◯	little	◯	sick	◯
awake	◯	fast	◯	long	◯	sleepy	◯
bad	◯	fine	◯	loud	◯	slow	◯
better	◯	first	◯	mad	◯	soft	◯
big	◯	full	◯	naughty	◯	sticky	◯
black	◯	gentle	◯	new	◯	stuck	◯
blue	◯	good	◯	nice	◯	thirsty	◯
broken	◯	green	◯	noisy	◯	tiny	◯
brown	◯	happy	◯	old	◯	tired	◯
careful	◯	hard	◯	orange	◯	wet	◯
clean	◯	heavy	◯	poor	◯	white	◯
cold	◯	high	◯	pretty	◯	windy	◯
cute	◯	hot	◯	quiet	◯	yellow	◯
dark	◯	hungry	◯	red	◯	yucky	◯
dirty	◯	hurt	◯	sad	◯		

16. *Words About Time* (12)

after ○	later ○	now ○	tomorrow ○
before ○	morning ○	time ○	tonight ○
day ○	night ○	today ○	yesterday ○

17. *Pronouns* (25)

he ○	me ○	their ○	we ○
her ○	mine ○	them ○	you ○
hers ○	my ○	these ○	your ○
him ○	myself ○	they ○	yourself ○
his ○	our ○	this ○	○
I ○	she ○	those ○	
it ○	that ○	us ○	

18. *Question Words* (7)

how ○	when ○	which ○	why ○
what ○	where ○	who ○	

19. *Prepositions and Locations* (26)

about ○	beside ○	next to ○	there ○
above ○	by ○	of ○	to ○
around ○	down ○	off ○	under ○
at ○	for ○	on ○	up ○
away ○	here ○	on top of ○	with ○
back ○	inside/in ○	out ○	
behind ○	into ○	over ○	

20. *Quantifiers and Articles* (17)

a ○	any ○	not ○	the ○
all ○	each ○	none ○	too ○
a lot ○	every ○	other ○	
an ○	more ○	same ○	
another ○	much ○	some ○	

Helping Verbs (21)

am	○	do	○	is	○	were	○
are	○	does	○	lemme/let me	○	will	○
be	○	don't	○	need/need to	○	would	○
can	○	gonna/going to	○	try/try to	○		
could	○	gotta/got to	○	wanna/want to	○		
did/did ya	○	hafta/have to	○	was	○		

Connecting Words (6)

and	○	but	○	so	○
because	○	if	○	then	○

How Children Use Words

	Not Yet	Sometimes	Often
Does your child ever talk about past events or people who are not present? For example, a child who saw a parade last week might later say parade, clown, or band.	○	○	○
Does your child ever talk about something that's going to happen in the future, for example, saying "choo choo" or "airplane" before you leave the house for a trip, or saying "swing" when you are going to the park?	○	○	○
Does your child talk about objects that are not present such as asking about a missing or absent toy, referring to a pet out of view, or asking about someone not present?	○	○	○
Does your child understand if you ask for something that is not in the room, for example, by going to the bedroom to get a teddy bear when you say "where's the bear?"	○	○	○
Does your child ever pick up or point to an object and name an absent person to whom the object belongs? For example, a child might point to mommy's shoe and say "mommy".	○	○	○

Part II: Sentences and Grammar

A. Word Endings/Part I

	Not Yet	Sometimes	Often
1. To talk about more than one thing, we add an "s" to many words. Examples include cars (for more than one car), shoes, dogs, and keys. Has your child begun to do this?	○	○	○
2. To talk about ownership, we add an "'s", for example, Daddy's key, kitty's dish, and baby's bottle. Has your child begun to do this?	○	○	○
3. To talk about activities, we sometimes add "ing" to verbs. Examples include looking, running, and crying. Has your child begun to do this?	○	○	○
4. To talk about things that happened in the past, we often add "ed" to the verb. Examples include kissed, opened, and pushed. Has your child begun to do this?	○	○	○

B. Word Forms

Following are some other words children learn. Please mark any of these words that your child uses.

Nouns

children	○	men	○	teeth	○
feet	○	mice	○		○

Verbs

ate	○	fell	○	made	○
blew	○	flew	○	ran	○
bought	○	got	○	sat	○
broke	○	had	○	saw	○
came	○	heard	○	took	○
drank	○	held	○	went	○
drove	○	lost	○		

C. Word Endings/Part 2

Young children often place the wrong endings on words. For example, a child might say "Auntie goed home." Mistakes like this are often a sign of progress in language. In the following lists, please mark all the mistakes of this kind you have heard your child say recently.

Nouns

blockses	○	foots	○	mouses	○	toeses	○
children	○	mans	○	shoeses	○	tooths	○
childs	○	mens	○	sockses	○		
feets	○	mices	○	teeths	○		

Verbs

ated	○	bringed	○	broked	○	doed	○
blewed	○	buyed	○	camed	○	dranked	○
blowed	○	breaked	○	comed	○	drinked	○

(continued)

eated	○	gotted	○	losted	○	satted	○
falled	○	haved	○	maked	○	sitted	○
flied	○	heared	○	ranned	○	taked	○
getted	○	holded	○	runned	○	wented	○
goed	○	losed	○	seed	○		

	Not Yet	Sometimes	Often
Has your child begun to combine words yet, such as "nother cracker", or "doggie bite?"	○	○	○

If you answered not yet, please stop here. If you answered sometimes or often, please continue.

D. Examples

Please list three of the longest sentences you have heard your child say recently.

1. ______________________________

2. ______________________________

3. ______________________________

E. Complexity

In each of the following pairs, please mark the one that sounds MOST like the way your child talks right now. If your child is saying sentences even more complicated than the two provided, just pick the second one.

1. Two shoe. ○
   Two shoes. ○
2. Two foot. ○
   Two feet. ○
3. Daddy car. ○
   Daddy's car. ○
4. (Talking about something happening right now)
   Kitty sleep. ○
   Kitty sleeping. ○
5. (Talking about something happening right now)
   I make tower. ○
   I making tower. ○
6. (Talking about something that already happened)
   I fall down. ○
   I fell down. ○
7. More cookie! ○
   More cookies! ○
8. These my tooth. ○
   These my teeth. ○
9. Baby blanket. ○
   Baby's blanket. ○
10. (Talking about something that already happened)
    Doggie kiss me. ○
    Doggie kissed me. ○
11. (Talking about something that already happened)
    Daddy pick me up. ○
    Daddy picked me up. ○
12. (Talking about something that already happened)
    Kitty go away. ○
    Kitty went away. ○
13. Doggie table. ○
    Doggie on table. ○
14. That my truck. ○
    That's my truck. ○
15. Baby crying. ○
    Baby is crying. ○
16. You fix it? ○
    Can you fix it? ○
17. Read me story, Mommy. ○
    Read me a story, Mommy. ○
18. No wash dolly. ○
    Don't wash dolly. ○
19. Want more juice. ○
    Want juice in there. ○
20. There a kitty. ○
    There's a kitty. ○
21. Go bye-bye. ○
    Wanna go bye-bye. ○
22. Where mommy go? ○
    Where did mommy go? ○
23. Coffee hot. ○
    That coffee hot. ○
24. I no do it. ○
    I can't do it. ○
25. I like read stories. ○
    I like to read stories. ○
26. Don't read book. ○
    Don't want you read that book. ○
27. Turn on light. ○
    Turn on the light so I can see. ○
28. I want that. ○
    I want that one you got. ○
29. Want cookies. ○
    Want cookies and milk. ○
30. Cookie mommy. ○
    Cookie for mommy. ○
31. Baby want eat. ○
    Baby want to eat. ○
32. Lookit me! ○
    Lookit me dancing! ○
33. Lookit! ○
    Lookit what I got! ○
34. Where's my dolly? ○
    Where's my dolly name Sam? ○
35. We made this. ○
    Me and Paul made this. ○
36. I sing song. ○
    I sing song for you. ○
37. Baby crying. ○
    Baby crying cuz she's sad. ○

Child Report Form
CDI: Words and Sentences

ID number ______ Date of report ______ Date CDI completed ______

Child's name ______ FIRST MIDDLE LAST Child's birth date ______

Parent/guardian ______ FIRST LAST Gender ____ Child's age in months ____

PART I: WORDS CHILDREN USE

Vocabulary Checklist

Words Produced: Number: ______ (of 680) Percentile: ______

How Children Use Words

			Percentage of affirmative answers at this child's age
Past:	☐ yes	☐ no	______
Future:	☐ yes	☐ no	______
Absent Object (Production):	☐ yes	☐ no	______
Absent Object (Comprehension):	☐ yes	☐ no	______
Absent Owner:	☐ yes	☐ no	______

PART II: SENTENCES AND GRAMMAR

Word Endings/Part 1

			Percent of affirmative answers at this child's age
Plural (-*s*):	☐ yes	☐ no	______
Possessive (-*'s*):	☐ yes	☐ no	______
Progressive (-*ing*):	☐ yes	☐ no	______
Past tense (-*ed*):	☐ yes	☐ no	______

Word Forms Number: ______ (of 25) Percentile: ______

Word Endings/Part 2 Number: ______ (of 45)

			Percent of affirmative answers at this child's age
Combining	☐ yes	☐ no	______

Examples

Length in morphemes of child's three longest sentences (M3L): 1. ______ 2. ______ 3. ______

M3L (mean): ______ Percentile: ______

Complexity

Number of times the more complex sentence is selected: ______ (of 37) Percentile: ______

DO NOT BREAK.
These forms are not sold individually. They are available only as a set. Resellers should not open shrink-wrapped sets.

ISBN-13: 978-1-55766-893-6
ISBN-10: 1-55766-893-0

· P A U L · H ·
BROOKES
PUBLISHING CO®
Post Office Box 10624
Baltimore, MD 21285-0624
www.brookespublishing.com

Words and Sentences pack includes 25 forms.

Manufactured in the U.S.A.

Appendix D.

Further variables selected by the second study's Variable Importance Plot.

- Aware 1 = The child is aware of the music being played
- Alone/Interacting 2 = The child is interacting with someone in listening to the music being played
- Multimod3 = The music is rendered in a multimodal context , engaging all of the child's senses.
- Av_dail_hr = Average daily hours mother spends with child
- Lyrics.1 = The music being played contains lyrics
- Trial = The phone call number (out of 21 allowed)
- Voc = The child's vocabulary scale score in the MacArthur-Bates Communicative Development Inventories.
- Fam_source1 = The source of the music being played is familiar to the child
- Age = Age of the child's participating parent
- Pr.pst2 = Music took place not during the phone call but before
- Mus_gen3 = Music being played corresponded to nursery rhymes
- Age_t_ch = Age of the child
- Mus_source2 = Music is being sung or performed by kin
- Hour_cat3 = Phone call took place between 14 and 18hrs in the evening
- Mus_source5 = Music is being played by a device

Appendix E.

Third study's questionnaire assessing participants' demographics and impressions concerning their children's development, as well as information about the participating child's family composition.

Music to my Ears! III

Dear Participant Thank you again for joining Music to my Ears! This questionnaire will ask for basic demographic information as well as a partial assessment of your toddler's language development. Please complete this questionnaire in one go, without closing the window until you are done. Thank you very much!

1. Your name

2. Target Child’s name

3. Name

4. Your age

5. Latest QualificationLevel 1 (first certificate; GCSE - grades 3, 2, 1 or grades D, E, F, G; national vocational qualification (NVQ), music grades 1, 2 and 3; or equivalent) Level 2 (CSE - grade 1; GCSE - grades 9, 8, 7, 6, 5, 4 or grades A*, A, B, C; intermediate apprenticeship; music grades 4 and 5 O level - grade A, B or C; or equivalent) Level 3 (A level, access to higher education diploma; advanced apprenticeship; international Baccalaureate diploma; tech level; or equivalent) Level 4 (certificate of higher education (CertHE); higher apprenticeship; higher national certificate (HNC); or equivalent) Level 5 (diploma of higher education (DipHE); foundation degree; higher national diploma (HND); or equivalent) Level 6 (degree apprenticeship; degree with(out) honours - for example bachelor of the arts (BA) hons, bachelor of science (BSc) hons; graduate certificate; or equivalent) Level 7 (integrated master’s degree, for example master of engineering (MEng); master’s degree, for example master of arts (MA), master of science (MSc); postgraduate certificate; or equivalent) Level 8 (doctorate, for example doctor of philosophy (PhD or DPhil); or equivalent)

Level 1

Level 2

Level 3

Level 4

Level 5

Level 6

Level 7

Level 8

None

6. Occupation (other than child care), if any

7. Number of hours per week invested in occupation (if none, please selelct "0")

Does not apply

1-5

6-10

11-15

16-20

21-25

26-30

31-35

36-40

More

8. Average daily hours spent with target child (sleeping does not count)

9. Your mother tongue

10. Language(s) spoken (by you or other primary caregivers) to target child and average proportion (e.g. English 70%, Italian 30%)

11. Number of Children

1
2
3
4
5
More (please specify number of children)

12. Age of Child(ren) in years and months (e.g. 1 year 10 months, 3 years 2 months)

13. How much do you agree with the following claims?(-3 = I do not agree at all, 3= I completely agree)

	-3	-2	-1	1	2	3
My child understands everything I say to her/him						
My child understands sentences (not just single words, but combinations of them)						
Sometimes, my child understands me, but pretends like she/he does not						
Sometimes my child does not cooperate due to the fact that he/she does not understand what I'm asking him/her						
My child has learned to walk						

14. How autonomous would you say your target child is?

Very dependent										Very autonomous
0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Appendix F.

The Language Use Inventory for Young Children

Language Use Inventory

INSTRUCTIONS

Complete By: April 11, 2014 8:30 pm

As a parent, the information you can provide about your child's communication across a wide variety of settings is unique and valuable.

Please read these instructions carefully before beginning to complete this questionnaire.

1. It is very important that ALL questions are answered. The system will not allow you to leave any questions unanswered.
2. Please complete the entire questionnaire in a single day if possible, or two at most. The system will expire the questionnaire after two days.
3. If your child speaks a language other than English at home, when answering the questions you should include what your child says in **ANY** language. For example, many questions will ask whether your child uses words for a particular purpose (e.g., to describe what he or she is currently doing); you should respond "yes" even if your child only does so in his or her non-English language.
4. You may consult with other people (e.g., spouse, grandmother, nanny, daycare teacher) about any items on the questionnaire should you find this helpful in deciding on the appropriate response.

PART 1

How your child communicates with gestures

These first two sections, **A and B**, will ask you about your child's use of **gestures**. If your child is not using a gesture described below anymore, but did use the gesture in the past, mark the box "not anymore." You will be asked more about your child's use of words later in the questionnaire.

A: HOW YOUR CHILD USES GESTURES TO ASK FOR SOMETHING

11 questions remaining ▲

At this time, does your child use any of the following gestures to ask you for something, with or without words?

	Never	Rarely	Sometimes	Often	Not Anymore
1) take your hand, push it, or lead you, to what he/she wants	○	○	○	○	○
2) put a toy or book in your lap, or climb into your lap with a toy	○	○	○	○	○
3) lift his/her arms to ask to be carried	○	○	○	○	○
4) hold up an object to show you what he/she wants (e.g., hold up a cup to ask for milk)	○	○	○	○	○
5) reach for or point at what he/she wants	○	○	○	○	○
6) get in a starting position so that you will play a game again (e.g., hold his/her feet up so that you will tickle them again)	○	○	○	○	○
7) look where something is that he/she wants you to get	○	○	○	○	○
8) look at something that he/she wants you to do something with	○	○	○	○	○
9) look at you when he/she wants information from you	○	○	○	○	○

For each item below, please mark the box that **best applies** to your child at this time:

10) my child tries to get my help using gestures	○	○	○	○	○
11) my child uses gestures to get me to play with him/her	○	○	○	○	○

B: HOW YOUR CHILD USES GESTURES TO GET YOU TO NOTICE SOMETHING

2 questions remaining ▲

If your child finds something that **interests** him/her, would he/she use any of the following **gestures**, with or without words?

	Yes	No
1) **point** at what he/she finds interesting		
2) **bring** to you, **show** to you, or **give** you something he/she finds interesting		

PART 2

Your child's communication with words

C: TYPES OF WORDS YOUR CHILD USES

21 questions remaining ▲

Has your child begun to **say** any of the following **types of words?**

Mark "yes" even if your child uses only one of the example words.

	Yes	No
1) people (e.g., mama/mommy or dada/daddy, baby)		
2) food items (e.g., juice, milk, cookie)		
3) animals (e.g., dog, kitty)		
4) body parts (e.g., eye, nose)		
5) vehicles (e.g., car, boat, train)		
6) toys (e.g., ball, block, doll)		
7) clothing (e.g., diaper, shoe, sock)		
8) household items (e.g., cup, spoon, bottle, light)		
9) "no" or "yes"		
10) "up," "down," "open" or "close"		
11) "in," "out," "on" or "off"		
12) "gone" or "all gone"		
13) "there" or "did it" when he/she has succeeded at something		
14) "here" or "there"		
15) "this" or "that"		
16) "go," "going" or "went" (e.g., Go away.; Doggie going.)		
17) "do," "doing" or "did" (e.g., Do it.; Did it.)		
18) "make," "making" or "made" (e.g., Making cookies.; Made that.)		
19) "get," "getting" or "got" (e.g., Get it.; Got cookies.)		

What were your child's first three words? (leave blank if you can't remember)

1) ______ 2) ______ 3) ______

For the items below, please mark the box that **best applies** to your child **at this time:**

	Never	Rarely	Sometimes	Often
20) it is fairly easy for me to teach my child a new word				
21) it is fairly easy for me to know when my child and I are both talking about the same thing				

D: YOUR CHILD'S REQUESTS FOR HELP

7 questions remaining

Does your child **ask for your help:**

	Yes	No
1) by using the word "help"		
2) by telling you what he/she wants by name (e.g., milk, cookie)		
3) by asking you to do something again (e.g., More.; Do it again.)		
4) to play a game		
5) by asking you to do something difficult (e.g., to open a door, to carry something heavy)		
6) by asking you to make a toy work, or to fix a toy		

For the item below, please mark the box that **best applies** to your child **at this time:**

	Never	Rarely	Sometimes	Often
7) My child uses his/her words to ask for my help				

E: YOUR CHILD'S INTERESTS

2 questions remaining

What are your child's three **favourite play activities?**

1)

2)

3)

	Yes	No
4) Does your child seem to be interested in things that you find **unusual or that other children of the same age are not interested in?**		

	Yes	No
5) Does your child seem to be **excessively** interested in **one** thing?		

PART 3

Your child's longer sentences

As you begin Part 3, please note that if your child is using only a few words, you will likely be answering "no" to many questions. However, it is very important that you **fill out ALL of part 3** as this will provide the best overall picture of your child's communicative ability.

First, as an estimate of **how long** your child's sentences currently are, please answer the following two questions:

	Never	Rarely	Sometimes	Often
Has your child begun to use sentences of **more than 2 words?**				
Has your child begun to use sentences of **more than 4 words?**				

F: HOW YOUR CHILD USES WORDS TO GET YOU TO NOTICE SOMETHING

6 questions remaining

Does your child ever try to **get your attention** by doing any of the following things?

	Yes	No
1) naming something he/she is interested in (e.g., Kitty!; Airplane!)		
2) asking you to "Look!" or "Watch me!"		
3) asking "Can I try?", "Can I do it?" or something similar		
4) saying "You know what?" or "Guess what?"		

For the items below, please mark the box that **best applies** to your child **at this time:**

	Never	Rarely	Sometimes	Often
5) my child uses words to ask me to look at him/her, or at what he/she **is doing**				
6) my child uses words to ask me to look at something he/she **is interested in**				

G: YOUR CHILD'S QUESTIONS AND COMMENTS ABOUT THINGS

9 questions remaining

When **talking about things** like toys, does your child ever talk about or ask about:

	Yes	No
1) **what** something is (e.g., What's this?; What's that?)		
2) **where** something is (e.g., Where's dolly?; Ball's in the box.)		
3) **more information** about something such as what it is used for (e.g., What's that for?)		
4) **why** something happened (e.g., Why did that car stop?)		
5) what something is **doing** (e.g., Car's going.)		
6) who something **belongs** to (e.g., Daddy's car.; Mine., Mommy's.)		
7) how something **tastes, sounds, feels or smells** (e.g., yummy, loud, soft, stinky)		
8) how something **looks or what he/she thinks of it** (e.g., its colour, shape; whether it's broken, pretty)		
9) how something is **similar** to something else (e.g., Just like Daddy's.)		

H: YOUR CHILD'S QUESTIONS AND COMMENTS ABOUT THEMSELVES OR OTHER PEOPLE

36 questions remaining

Which of the following things have you heard your child talk about?
Note: It's okay if your child does not use "I" or uses his/her own name or "me" instead of "I" in these examples.

	Yes	No
1) what his/her own **name** is (e.g., My name's Alicia.; I'm Brendan.)		
2) **who** someone is or what their **name** is (e.g., Who's that?; What's your name?)		
3) **where** he/she is (e.g., I'm in here.)		

3) **where** he/she is (e.g., I'm in here.)

4) **where** someone else is (e.g., Where's Daddy?; Mommy's here.)

5) what he/she is **doing** (e.g., I'm helping Mommy.)

6) what another person is **doing** (e.g., Baby's sleeping.)

7) what he/she **wants or doesn't want** (e.g., I want ice cream.; I don't want it.)

8) what someone else **wants or doesn't want** (e.g., Ben wants the truck.)

9) whether he/she **likes** or **dislikes** something (e.g., I don' t like apples.)

10) what someone else **likes** or **dislikes** (e.g., Do you like carrots?; Daddy likes ice cream.)

11) say **how old** he/she is (e.g., I'm three.)

12) ask someone **how old** they are (e.g., How old are you?; Are you six?)

13) how he/she is **feeling physically** (e.g., tired, cold, thirsty, sick, hungry)

14) how someone else is **feeling physically** (e.g., Mommy sick?)

15) how he/she is **behaving** (e.g., silly, nice, bad)

16) how someone else is **behaving** (e.g., Jamie's being mean.; That boy's nice.)

17) what he/she **thinks of something** (e.g., pretty boat, nice pictures, yucky broccoli, good cookies)

18) what someone else **thinks of something** (e.g., Daddy thinks broccoli is yucky.)

19) what he/she **wants** or **has to do** (e.g., I want to play.; I have to put shoe on.)

20) what someone else **wants** or **has to do** (e.g., Mommy wants to sleep.)

21) what he/she is going to do (e.g., I'm gonna draw a house.)

22) what someone else is **going to do** (e.g., Daddy's gonna buy me an ice cream.)

23) how he/she **feels emotionally** (e.g., sad, happy, angry)

24) how someone else **feels emotionally** (e.g., Baby sad?)

25) **why** someone **feels** the way they do (e.g., Why are you sad Mommy?)

26) that he/she wants to do something on **his/her own** (e.g., I want to do it.; Me do it.)

27) how he/she **can or can't do something** (e.g., I can run fast.; I can't draw a dog.)

28) how someone else **can or can't do something** (e.g., You can't see me.; You can't do it?)

29) ask someone **how** they **did something** (e.g., How'd you do that?)

30) ask **why** someone is **doing or did** something (e.g., Why's that boy crying?)

31) ask someone **why** they **won't do** something (e.g., Why won't you play with us?)

32) ask more detailed questions about people's **lifestyles** (e.g., Do you have a bike?; Do you live here?)

For each item below, please mark the box that **best applies** to your child at this time:

	Never	Rarely	Sometimes	Often
33) my child makes comments or asks about **objects**				
34) my child makes comments or asks about **people**				
35) my child's questions and comments are usually **appropriate** and **relevant** (not strange or out-of-place)				
36) my child uses language in a **spontaneous** and **natural** way that does not seem mechanical, memorized, or part of a routine				

I: YOUR CHILD'S USE OF WORDS IN ACTIVITIES WITH OTHERS

14 questions remaining

Does your child do any of the following?

		Yes	No
1)	ask an adult to show him/her how to do something		
2)	like to show other people how to do something		

If your child were **playing a game** such as rolling a ball down a slide with you or another child, would your child do any of the following things?

		Yes	No
3)	describe what he/she is doing (e.g., I'm eating.; I'm getting the ball.)		
4)	describe what another person in the game is doing (e.g., Mommy's next.; You dropped it.)		
5)	repeat something the other person said (e.g., Down it goes.)		
6)	tell another person what to do in the game (e.g., Do it again.; Wait!)		
7)	tell another person to stop doing something (e.g., Don't do that.; Stop!)		
8)	describe something they are doing with someone else (e.g., We're jumping.)		
9)	ask for a turn (e.g., My turn now.)		
10)	ask another person in the game about something (e.g., Is that your ball?; My turn?)		

Does your child talk with you, a brother or sister, or playmate about any of the following things?

		Yes	No
11)	toys		
12)	TV, movies, video or computer games		
13)	games to play		
14)	rules		

J: TEASING AND YOUR CHILD'S SENSE OF HUMOUR

5 questions remaining

Does your child **laugh or try to make others laugh** by doing any of the following things?

		Yes	No
1)	**saying wrong things** in a teasing way (e.g., giving the wrong name for something even though you know he/she knows the right name for it)		
2)	teasing others by calling them silly names (e.g., You're silly.; You're poopy.)		
3)	**doing something wrong** in a teasing way (e.g., putting puzzle pieces in the wrong place even though you know he/she knows how to do the puzzle)		
4)	making up silly rhymes		
5)	telling jokes		

K: YOUR CHILD'S INTEREST IN WORDS AND LANGUAGE

12 questions remaining

Have you noticed that your child does any of the following things?

		Yes	No
1)	answers questions that you ask while reading books		
2)	imitates words or phrases you say or that he/she has heard on TV or video		
3)	plays with the pronunciation of words (e.g., tries saying words different ways, rhymes)		

		Yes	No
4)	answers "What colour?" questions with a colour name (colour name doesn't have to be correct)		
5)	answers "How old are you?" or "How many?" with a number (number doesn't have to be correct)		
6)	likes to count or point as someone else is counting		
7)	during pretend play, he/she makes the dolls or animals talk to each other		
8)	talks about what other people said (e.g., My mommy said...)		
9)	asks about the meaning of words that are new for him or her (e.g., What's a caterpillar?)		
10)	is interested in logos and the writing on toys and objects such as store signs or billboards		
11)	rehearses talk for future interactions, such as meeting new children		
12)	asks to be told a familiar story about a family event (e.g., the day he/she was born)		

L: YOUR CHILD'S INTEREST WHEN TALKING

4 questions remaining

		Yes	No
1)	Does your child talk about some things that you find **unusual**?		
2)	Does your child seem to talk **only about one topic excessively**?		
3)	When your child talks, does it seem like he/she is often just **repeating** word-for-word what he/she has heard without really understanding what it means?		
4)	Does your child ever **make up new words** that you find **interesting** or **out-of-the-ordinary** (e.g., making up the name "bumblenest" for "beehive")?		

5) What would you say are the **three** things your child **talks about most?**

1.

2.

3.

M: HOW YOUR CHILD ADAPTS CONVERSATION TO OTHER PEOPLE

15 questions remaining

		Yes	No
1)	If you ask your child a question, does he/she usually stay on the topic and try to answer as best as he/she can?		
2)	If your child doesn't understand something you have said to him/her, does he/she usually say something like "Huh?", "What?" or "What did you say?" to try to better understand you?		
3)	If you said "Give me that one," and your child was not sure which one you wanted, would he/she try to make sure which one you wanted asking you a question like "This one?"		
4)	When listening to a story, does your child ask relevant questions or make relevant comments?		
5)	If you are talking with someone else and your child is nearby, does your child sometimes join in with a comment related to what you are talking about?		

Suppose you and your child had spent the day at the zoo, and that evening Grandma (or someone else in the family) was interested in what happened. Could your child:

Yes No

		Yes	No
6)	tell Grandma about the zoo if given prompting questions such as "What did you see at the zoo today?"		
7)	tell Grandma about it spontaneously, without needing much adult help or prompting		

Does your child talk about past events in any of the following ways?

		Yes	No
8)	he/she will mention something that just happened (e.g., My dolly broke.; Daddy spilled it.)		
9)	he/she will try to answer when asked to tell someone about something (such as when asked "Tell Daddy what you saw today.")		
10)	he/she will try to answer when you ask "Do you remember ..."		

Suppose you came home and hadn't seen your child all day. Would he/she:

		Yes	No
11)	say something about what he/she is currently doing (e.g., I'm making cookies!)		
12)	spontaneously tell you about an earlier event of that day, that you did not know about		

Does your child ever use the word "know" or "think" in any of the following ways?

		Yes	No
13)	says "You know what?" before telling you something		
14)	states that he/she is certain by using "know" (e.g., I know that's a hamster.)		
15)	uses "think" when he/she is not sure (e.g., I think it's in the drawer.)		

N: HOW YOUR CHILD IS BUILDING LONGER SENTENCES AND STORIES

36 questions remaining ▲

Please mark any of these **words** that your child has begun to use:

		Yes	No			Yes	No			Yes	No
1)	wish			8)	might			15)	possibly		
2)	hope			9)	could			16)	perhaps		
3)	forgot			10)	can			17)	after		
4)	think			11)	would			18)	going to (gonna)		
5)	know			12)	will			19)	before		
6)	remember			13)	maybe			20)	later		
7)	must			14)	if			21)	want to (wanna)		

			Yes	No
22)	and	(e.g., We saw trains **and** planes **and** trucks.)		
23)	then	(e.g., ...and **then** we saw rabbits.)		
24)	because	(e.g., I'll help you, **'cause** I'm the fireman.)		
25)	so	(e.g., It's not cooked yet, **so** it has to go in the oven.)		
26)	but	(e.g., Now I'm big, **but** I used to cry.)		
27)	well	(e.g., **Well**, I think it's here.)		
28)	just	(e.g., I'm **just** taking it for a little while.; I'm **just** helping.)		

			Yes	No
29)	next	(e.g., **Next**, we saw bears.)		
30)	when	(e.g., **When** it's night, I go to bed.)		
31)	actually	(e.g., **Actually**, I don't like tomatoes.)		

When your child tells you a story, or part of a story:

		Yes	No
32)	can you follow **who the people are** in the story?		
33)	can you usually follow **what is happening** in the story?		
34)	can your child **link the events** in the story in a way that makes sense?		
35)	can your child **change the topic** in a way that doesn't leave you confused?		
36)	does he/she sometimes use words such as **"today," "yesterday,"** or **"tomorrow"**?		

SAVE AND EXIT SUBMIT

Appendix G. Focus group transcription

Wolfon College, Plommer House, 12March 2017

[Researcher briefly introduces the topic]

Researcher: Why do you use it [IDSp] when you feel like using it?

DorothyF: usually to keep him [Richie] calm. If he's agitated, slowing down the speech and the rhythm seems to calm him down as well because you can just… get back to a rhythm that seems more soothing [I1 S1 E1 EM1]. But I think it also sooths me, selfishly enough (chuckles). Because when he's getting agitated you don't wanna get agitated too, so if you force yourself into that frame of mind, that slowness, it actually calms both of us down, in a weird way. [I2 S2 E2 SS1 EM2]

DanielS: whenever he's having a tantrum [T1] and he's really agitated, if we get upset and tell him off, and afterwards he'd keep saying 'are you happy? are you happy?' and he can almost only calm down if you're happy, then he gets happy. [EmC1 EM2]

DF: so, it is in that sense… the worst thing if he's upset is for you to be upset and stressed, which is hard if you're having a tantrum [T2] in the middle of John Lewis [AE1] [laughs] but you have to get to that point

DS: he won't be calm unless you are calm [EC2 S3 Mnt1]

Paula: it strikes me that I spoke in that way when I was trying to engage in eye contact [EyC1] with him [her son, Newton] [NP1]. So I think that was definitively powerful for him. Talking in that way to really sort of engage them [En1], whether I was trying to sooth them or not. It was very good at turning into that particular closeness [AP7 In1], whether I was looking for their attention [En2 NP2], or trying to distract them, it was all very sort of a close moment.

Grace: what I was thinking was that when you want them to understand something [En3 CC1 NP3], you use also a special kind of tone like when trying to explain the way something works [CL1 NP4].

Ava: I was thinking you don't just use it when you're trying to calm down, but also when you're trying to get them to behave or not-do something they're trying to do, as in 'hold my hand' or 'don't run into the road' [ICC D1] (…) you change your tone when you're trying to get some control [LP1], you can change your tone to be slightly more authoritative or… [Au1 ICC1]

R: authoritative *and still* infant-directed tone of voice?

A: don't know, I'm still trying to figure it out (laughs)

DF: no, but I think some of the value of IDSp obviously is… there are ways in which the vocabulary is not as well understood [LB1] so sometimes the IDSp is just to convey a safe space [AP1 PS1], engaging… like sitting at the table and asking 'how was your day?' and that sort of thing… I think the tone very important so that you can distinguish it from a tone that is more urgent when you need to use that in occasions [TS] [CS1]. So, in a way, if my speech with him would the same as I use with the people I work with it may not be…

[Richie suddenly cries out loud outside the focus group room and we all turn our heads to pay to attention to it]

DF: but I think it's that sense of getting a variety of messages when IDSp becomes an important messaging tool when not all of the vocabulary is understood [CC NP5]. And that's why it distinguishes so sizably… [CS2] you need to distinguish from when you have to say something like 'don't play in traffic' [D2] which is kind of a basic message and they need to learn that one too, and if it's delivered in the same way then it's harder to get that message across [NP6]. So, in a way the range has to be wider for children [TS2], which is why the speech has to be safer and slower [AP2 PS2], like the one you would use with your pet. I sometimes ask myself 'am I talking to my son like a dog?', like 'come here!'

[R comments on the fact that Secondary Babytalk lead to the idea of Attachment Vocalizations]

DF: I think for us, since we adopted Richie, we didn't have that same biological attachment [A1] when he was born; he didn't recognise the sound of our voices. We weren't there until two months later. I think the speech then became very much for us about providing that sense of safety for him [AP3 PS3]. Because he had travel from her mother for a few days, to a foster mother, then us, in the course of a two-months lifespan. And I still remember that first weekend that we had him; he didn't cry! Like, the whole weekend, he didn't cry. We didn't feed him— I'm laughing at this now — because he didn't cry for any food! And it was this sense of apprehension. I think the speech became very important in getting him to feel safety [AP4 PS4]. So it was very much about that kind of engagement. [En4]

[DF leaves the room to attend to Richie]

R: What about your son, Grace?

G: actually he entered the nursery when he was 15 months, so he's actually already starting to say little things [LA1], even when you don't know exactly what it was, he's trying to communicate. So, we're mostly trying to communicate with him in Spanish but, he's already improved a lot his sentences since he entered the nursery, and now he's using a lot of words, he can really express himself now with words. It's not with complete phrases though [LA2].

R: [to all] Do you think language acquisition had or has an impact in the amount or quality of the IDSp you're using?

Julia: so, if I'm addressing her, I'm teaching her how to speak, but I think I'm addressing her as a normal person. [H1]

R: to J, is she putting together syntactical phrases?

J: no…

R: [to DF and DS] how about you guys?

DS: I actually try not to do the babytalk, and try to talk to him like anybody else, you know, like to help him with the language. So I just talk to him normally, with sentences. [H2]

R: what I meant to say is that you mentioned that the register would sometimes compensate for the lack of understanding of words. Has that changed now that Richie actually understands words?

DF: yeah, I guess it has. I'm probably more relaxed in the sense of not thinking to switch [CS3 CC2 Ir1] the tone and using my normal voice because he understands what I'm saying in terms of the words. [LB2] So I think that makes it a bit easier to just communicate on a day to day basis. I'm from NY, I talk fast, I talk loud (laughs), so slowing myself down is sometimes hard, but, yeah, as he's gained more language, I guess it's less infant-directed and we're starting to have more conversations, so when there's a conversational element, then it becomes much more of a reflexive tone in terms of what's the tone he's using and how is that going, rather than something in which I'm trying to define the tone of the conversation [E2]. Because in the beginning it's not a conversation, they'll say a couple of words to you, but now it's a bit more conversational, so it becomes easier not to control your tone [CC3].

(…)

DF: there are still times when you use it. I think Grace, you were saying that when you want them to bring something on board, like when you're playing something and all of the sudden he realizes [CL2] 'oh, this goes this way', I can see myself falling back into it [IRP1], and I can see that kind of tone emerging again. But day to day, you know, sitting at the dinner table 'what did you do today?' is probably more my conversational tone than the tone I used when he was younger.

R: in which other stances you still use IDSp with Richie?

DF: there are these moments like 'he's learning something big right now' [CL3] and you can feel it [Emp1], and you can kind of come to it [Mnt2]. And I think the other big thing is during a tantrum [T3], during which IDsp becomes very important to keep him calm [EM3 IRP2]. So I think those are the instances in which I use it the most. And occasionally it becomes—embarrassingly enough —the voice you use in public [AE2 IRP3] then you really want him to do something and you don't wanna be embarrassed, like to get him to do something; honey, wouldn't it be nice to…'. So that's a third one, but I think it comes down more to skill acquisition [CL4] and soothing [S4] more than anything else now.

R: does any of this resonates with you, Paula or Sue?

Sue: I think something similar happened to me… whenever I would try to sooth her [S5] or to help processing something [EP1 Mnt3], that is when it came back [IRP4], but I particularly think with my son, I think I would also use it during tantrums [T4].

R: Xavier, would you like to add anything?

Xavier: I think I usually talk to her just like any other person [H3], but it's true that when she's doing something he shouldn't be doing, it's like 'OK, sweetie, you have to calm down'; you address her with a more relaxed tone, if you want her to relax [E3 EC3 EM4]

[several participants nod in agreement]

DS: there are many instances when you make your voice sound like that, and they will follow…

P: it so strikes me that you're trying to share them being in the same wave-length as you [E4], kind of bearing in mind what would be beneficial if they were… But you know that if they're becoming very aroused, and you want them to get back to the baseline [EM5], so it's like you take a deep breath, and you feel calm [SS2] and then by speaking to them… I think it's more than speech, it's a lot to do with your household-self [MM1], but speech is part of it, and that helps to regulate them back to that level that you hope it's the same same wave-length that you're on… [E5 EC4] hopefully (laughs)

G: it's not only when you want to sooth them, sometimes he's excited because he just discovered something and you want to actually follow him [CL4] in that case, so I don't want him to come down but I want to be like at his same level [H4].

R: what about you, Coline, have you any thoughts regarding this topic?

Coline: Yes, I remember… having commented on this. It could have been during two weeks that she started putting together these sentences. Perhaps I realized over a month… but it's like a quick switch [QC3], it's like they suddenly start making sentences [LA3].
It's worked like this in English too. My husband and I are trying to teach her English, and she suddenly started saying 'dog', 'water', etc. I feel it's that same switch. And yes, it's from one week to the next one. It's like they incorporate something and then just keep using it. Like a quick, drastic change [QC4].

R: and the fact that your daughter talks and you have at some point acknowledged it, has had any impact in the way you address them?

C: maybe yes. I think the fact that she understands makes you shift from child-like to more normal [LB4 SoS3].

J: Still, one keeps using childish expressions like 'oh look', but I do feel like the fact that they understand you means that you can address them more normally [NP7], with the tone you would use when addressing another person.

C: Before, you didn't know whether she was getting was you were transmitting or not [LB5], hence the slower, affectionate, and fragile or delicate tone [NP8]. I can now say 'Sweaty, put on your jumper because you're going to play outside'. Before she didn't understand, so you'd try to make up a game [LB6 CC3], so that she would help you, you can make up a song that she gets through the quality of the voice.

R: I see. You mentioned something more like 'fragile' in her that made you feel like talking to her like that. What do you think would happen if you would address her like any other adult, like I'm talking to you now? What would be the difference?

C: maybe none… (…) but I think that… I don't know why one does a change. Maybe like when you talk to an elderly person and you try to be a bit more formal… [SoS1 CS2]

R: like someone who has a different status?

A: yes, a different status, more serious, who you can't talk to… I think it's not something you think, because they're smaller [SoS2 CS3], and not just with my boy, but with other children, you make that shift in the voice…

R: do you feel as excited elsewhere as you would with your children? [explains disinhibited quality of IDSp]

G: I try to. I mean, obviously [SR] you're not gonna show your excitement like you're a kid [AI1], but when something makes me passionate I try to show excitement.

P: I think there is something not as inhibiting with your own child [IDD1], as there is perhaps with others [Des1, AI2]. So there might be some news that I share with other adults that I'm excited about, but that level of interaction with your own child is something that you can sort of just be yourself… [PP1 IDD2 PS5]

G: exactly [AI3 IDD3]

P: because, they love you however you are [UL1 A2] and you can equally be silly, or enthusiastic, or whatever it is because they love you whatever you're like so, you get that chance to be no inhibitions. [IDD4]

DF: definitively there is a joy that you can express [PEP1] that is just pure [PP2 ECx1]. It's just purer when you're with your child than when you're laughing at a joke at work or with your friends, it doesn't have the same innocence, in a weird way. And as such, it doesn't have the same inhibition [IDD 6]. There's no thought for it [Ir2], you know. And I think that's really nice so I think there's a degree of expressiveness that is much broader on the silly, fun side [PEP2] with him, than on the serious, frustration and all of that other stuff…

R: [to G, DS and A and M]: anything to add guys?

G: but, for example, now that I get more clearly the example in my head… I'm a very open person, I'm a very sharable person, so I try to express myself, actually, but my husband is exactly the opposite. So, that's why I see the difference between the way he talks to him and the way I talk to him, which is exactly the same outside actually. When he talks to people he tries to be really serious [AE3 SR2?], really like 'I'm an important person' and I'm more like 'yeah….'

A: I agree… [nodding]

Martha: I think I can relate to that. There are many things I would probably express more at work than in front of my daughter… hopefully (laughs out loud). But I agree, I feel the joy spectrum is indeed much purer [PEP3 PP3 ECx2], and more uninhibited than the other extreme of options.

R: I guess I was inaccurate before; it is not just any emotion what finds disinhibition in IDSp

[to DF]: what do you mean when you say 'pure'?

DF: I think very rarely an emotion in the outside world is that simple [ECx3]. You know, you might laugh at a joke but at the same time feel guilty that you laughed at a co-worker and you're being silly [SR3]. When you laugh with your child about whatever there's nothing else in the outside of that, it's just the laugh [DD1]. It's simpler in a weird way, it's not…

DS: there's no 'but'

DF: there's no 'but', there's no layers of the onion of complexity, it's just pure [ECx4]. And there are moments in the world that you get to feel that purity, but they're really rare compared to when you're home with the child [DD2]. It's very different. And I think it's that purity of it that can mean to the fact that the intensity feels much different, because there's nothing dampening it [ECx5].

R: what kind of things would dampen it?

DF: in the real world it's the normal things…

DS: the rules of society [SR4]

J: the rules of society! Like yeah, this is really funny, but maybe it's at somebody's expense. Sometimes when I get home and my daughter's already there, she'll find me in the kitchen, and he'll run at me at full speed, and you have to catch her, and he'll say 'nice catch mummy', and to me there's nothing better than that. So there're anything else talking at the back of your mind [ECx6 Ir3].

P: I was just thinking that it seems like when your son is laughing it seems like there's just the two of you [DD3], like there's nobody else in the world. But equally, I think when my son smiles and he looks at me, it's almost like, 'oh but you're laughing with me', like there isn't anything else. That strikes me as 'pure'.

DS: it feels almost like exactly the same emotion at the same time… [EC5 Emp]

P: yes, exactly!

R: can somebody elaborate more on the 'authoritative' instances?

Xavier: sure, right this morning, there was something dangerous my daughter seemed to be trying to do, so there was a moment when I said 'OK, I've asked you twice nicely… now I mean it'… so it's like… I…

S: Also, you see them smaller, more defenceless [PT1], so yes… you do that tone shift so that they can understand you [CC1] or look at you [EyC1].
But it has also happened to me, in a different sense, to be with her, playing, and changing my tone of voice when I want to make her see that there is something she shouldn't do [Au2 SR1 STT1 I3], or there is something dangerous [S2], or correct her [I4], I do use a slightly firmer tone of voice [TS2].

JP: still directed at her?

S: still directed at her. And I feel she does sense it [EC2 AP2]. That change in the voice from sweeter and more child-like, to one that conveys that there is something that shouldn't be done [Au3 SR2], or that she must know that care should be taken… [D2] you use a firmer tone, always with gentleness [STT2], but I do think they get that tone shift, that shift in the voice or in the intention of shat you're trying to say [AP3].

[R tries to recapitulate when Richie starts running and giggling loudly]

P: I was thinking I was tickling my son last night, and I realized that even at ten, my voice had gone mild when I was talking to him, when I was tickling him… so I was thinking that even though I thought it would wane, it would disappear, it doesn't… it's just certain occasions when it comes back I think… [IRP5]

[R explains an impression of his own; the fact that the introduction of language clarified the emergence of a will that is different from not understanding language]

X: So, in the example, I know she understands me, and he's completely ignoring me or…

DF: so, I think, in those cases, you know, there does have to be a way of appealing… the infant-directed, the soothing, the bringing things down a level in a way, and the authoritative edge, like 'I'm serious now' and it's hard because he's cute [PT1 ICC2] [everybody laughs] so there is that sense, almost back to the beginning, that the IDSp is off also to get you back to that mode… I guess for me IDSp is also about controlling your own… self-regulating, I guess. [SS3 EM6]
So, when the will comes out, then you have to have more than just two tones, you have to have a battery of tones, so that they can understand. [CS4 CC4 TS3]

R: like a pallete?

DS: reminds me of something you said in Switzerland. We got a house, we were dealing with the builders, in French, and you said you only had two tones; nice and shrill

DF: shrill! That's it! I can't deal do 'subtly obnoxious' and 'little bit frustrated' in a foreign language, so I thought that with every little problem the moment I got even remotely annoyed would escalate to the hugest problem in the world because in a foreign language I didn't have… the palette— as you called it —of tones, whereas with Richie at this point, you need it, because he is everywhere and going and you need to be able to regulate that through the tone of your voice and the words that you choose.

R: Ava, at which state are you in this regard?

A: I was just realizing that although he's tiny sometimes he just doesn't do something just because he decided not to. I think it's a mixture of things, not just IDSp. Sometimes you may try a simple sentence, and if he doesn't understand, you may change the tone [CCX Mnt4 NP9]

R: In this regard, is your children's use of language a confirmation of their understanding, or is understanding already there without verbal confirmation? Does language also open a space for more 'adult-like' emotions and not just the joyful ones?

G: exactly [the latter]. For instance, the 'yes' and 'no' he understands perfectly. So if I ask 'do you want a yogurt?' he's really frustrated because he's trying to tell me what he wants but he can't [LB3]. So, I ask him 'is it a banana you want?' and he would shake his head, until I ask him what he finally wants, like the milk, and he finally nods in agreement and you give him the milk and he's happy [Mnt5]. But before that he was so frustrated because he couldn't express himself, but I know he's actually understanding what I'm saying [Mnt6].

R: Is this related to, for instance, to the 'qualitative status shift' Coline previously referred to?

DF: no, I think it's true, I think the will comes up before the language, as Grace was saying, I think the will you can start to see before, but I thought the will before the language was like almost tentative, in the sense that he was trying to see 'could I assert myself?'. And once the language comes, he's much more confident in asserting himself [LA1 LP2] (laughs), so you can see, in a way, it's not just the 'adult emotions' I think it's the growing confidence of his control over his own environment [LP3]. You know, the 'I'm hungry, `I'm hungry!' and me 'what do you want?' and he can't answer the question… to 'I'm hungry. Yogurt, now!'. That sort of thing. And I think the thing that opens the door… the will is the first to appear, and then the language, but the latter ads confidence to the former, and I think that's what starts creating more complex emotions. He likes to tease you, in a sense of… you know, he'll pretend that he's really hurt, and cry… and he'll go 'haha, I got you'. That sort of thing suggests that complexity of manipulation comes with language [LP4 LA2 ICC3]

R: when did that start happening?

DS: I think just few months ago. You know, if he sees us kissing, he'll just stare like this [pouts and his shoulders drop], then he'll laugh naughtily like 'I got you'

DF: you know, it's that sense of increased confidence in his control of the environment [LP5], that he can do things to get the reaction that he wants. And he'll assess to what degree he can manipulate you… [ICC3]but then he figures it out and comes back down to…

C: it's true that once they've learned to talk you demand more from them, but at the same time, they still need much attention from one. But at the same time, they would like to do things on their own. Alondra is at a stage in which she likes doing things on her own, but still needs much affection. For instance, I ask her to tidy up. One expects them to cooperate, but at the same time when asking for something, there is a certain rejection, and a clash [ICC4]. You expect to be understood and helped, but at the same time there is a quarrel in which she is not responding. It's not because she doesn't understand, but because she doesn't want to. I think it's a natural phase of rebelliousness.
It's tricky. Sometimes you have to be firmer [Au4].

R: so since she talks, is there more conflict?

C: yes, absolutely. She understands, and understands that I understand. We both know we understand each other. But she might not want to cooperate.

X: [chuckles] Yeah, one starts by saying things warmly, because you think that maybe that way she'll better understand. But as time goes by, and you realize your child ignores you, you have to make a choice. You have to try something else [ICC4]. You may try something still childish, like inventing a game.

R: so this was after the second year…

X: yep, after the second year

R: and when did he started putting sentences together? Not necessarily full, but like having a syntactical sense?

DF: I think it was around the second year, so the will came first, but this capacity to manipulate and language came more or less together [LA4 LP6 ICC5]

G: my son is already doing that somehow, he can start some fake crying, or pretend not to be able to do something so that he gets someone else to do it for him [ICC6]

DF: yeah, I guess the fake crying came before language

M: in my case, although she cannot yet put a sentence together, she would go [bumbles babbling in a speech-like manner] and at the and say 'foot', so that you know her foot is involved in something else [Mnt7]

DF: also humour emerged only lately, with language [ECx7 LA5]

R: do the introduction of language and the waning of IDSp imply that children sometimes ask for it?

DS: I think he does, in a way… This morning he came to me and gave me a huge hug and said 'aw, my hero', and he said it in a babytalk way, so maybe he was asking for it, and I did reply in that sort of tone [IRD1]

G: my son is quite independent, but when he's tired or bored, he will call me by his side 'come, mom, come', and take my hand and ask me to sit, so that you know that he really wants you to stay with him, sit (…) when he really needs it, he knows how to ask for that kind of speech [IRD2]

R: but isn't it simply about getting attention? Or do you feel he also requires de IDsp?

G: yeah, yeah, because she tells you 'come, mummy' and she starts telling you something that will lead to that kind of talking

DF: I think it's… it's exactly that, it's more than just attention, there are times when he wants that soothing-ness [IRD3], and you can tell when it's not just 'I don't want you to pay attention to something else' but it's that soothing-ness that he needs [Mnt8]. And to DS's point, he'll bring out a bit of a baby voice, or bring out more of a [pout] [IRD4]

R: and in which kind of instances would they need you like that?

G: when I realise it more is when he's tired, when he wants to go to sleep. So he's like 'come one, mummy', and he knows that when we're at the sofa it's like 'relaxing time', and he's knows it [IRD5 Ment8]

DF: I get it, directed at me, at least, when there's a change in the routine [IRD6 A3]. For instance, I usually wake him up and then come back home at 6, but last week I had to stay in London Thursday night, and when I got back he was clingy for the whole weekend. The power of routine at this age is so huge, and the moment I lose any of my 'assigned roles' in that routine [laughs] he becomes much more focused on the soothing than anything else [IRD7].

R: just let me come back to the 'ID-authoritative tone'. Why not using a 'normal' authoritative tone? What difference does it make?

DS: I know he understands me, but I want to put on a little extra bit of 'and I mean it too' [STT1] if you like. I think just conveys that little extra bit of meaning I guess…

DF: and safely, because… I mean, there are many situations when things have escalated to a very authoritative tone, but it terrifies him. I mean, I terrifies him! [AP5 PS6 F2]

G: yeah, it's about not getting him scared [PS7 F3]

DS: that's actually the about trying to… because you have a range [TS4] from 'I'm asking nicely', to 'I'm asking a bit more urgently', to 'we really need to do it now', and the top one is… once he nearly run down into the road and I screamed 'no!' at the top of my lungs and it scared the crap out of him… but he stopped. [D3 A2 F4 LP7 PS7]

DF: you have to use that one so sparingly for it to mean anything [TS5]. Cause they're still young enough that although they're attached, there's still this fear… [A4 F5] there's this fear inside them and the IDSp with the authoritative tone embedded in it it's meant to take out the fear [AP6 PS8 PEP4 STT2].

G: My thoughts exactly. For example, when me and my husband are disagreeing and it escalates, my son knows perfectly the point when we are actually arguing and he cries, he shouts like 'stop!'. So if you're trying to be authoritative, you have to use a tone of voice that doesn't get over that top [PS9 TS6 STT3].

Sue: yes, there is an influence, and I think they get it [EC1 AC1]. I understand what you say, I may be talking to my husband, and then when addressing the child, one changes the tone [CS1]. If I was having an argument with my husband, when turning around and seeing your child, yes, you don't keep the same tone, you change it. First, I think you don't want to convey that there is a problem here [PS1], or because it's not an adequate tone for them [TS1].

R: so this seems to bring again the self-regulating aspect of it?

DF and G: yeah

DF: as Daniel mentioned before, if Richie gets scared, he'll start asking you 'are you alright?'. The anger makes them scared, even if it's just a word [F6].
IDSp is part of maintaining that safe environment. Language becomes a way of creating a safety net around you, for both of us actually [PS10 SS4 STT4].

R: since you've established that children that already talk will keep looking for moments of close encounter and so on, has music been a resource to you in anyway?

DF: we've had different songs at different phases [M1 MPP1]. Right now it's the song from Trolls. He wasn't that interested in the movie, all he wants is the song. But when he was smaller, it was the Muppets, then it was Frozen… [MP1]

DS: when he was really little, like 3 or 4 months old, there was one piece I could put on and he'd be asleep like in 3 mins. It was one of Haydn's symphonies. But only that one. He knew it straight away and [snores] it was like an off-switch [M2 A5 S6 MP2]

R: so he had like a 'favourite song' from very early on…

DF: so there were a number of songs in the sense of… music that we could put on that would change the mood… communicate something [M3 EM7]. Right now Trolls is just a dancing thing… [Da1]

DS: yeah, but it's something between you and him… [DD4]

P: as you were saying this I was thinking that we sort of moved from a stage where we had singing classes and we sung to the children and there were many pieces of music that we enjoyed over the years [MPP2]. And then we had a piece that I knew I played it when I was pregnant and I knew that my children recognized it every time… [PM1] but now it's dancing [Da2]. The music is still involved very much, but it's just that way of sort unwinding down at the end of the day… you know, if you're a bit tired or it's been a hectic day or it's been a bit stressful, we dance in the kitchen [7]. I don't even care if people could see me… [Des2]

R: well, we talked about this sort of 'safe space'…

G: as I said, he prefers French music over the lullabies at the nursery because that's the music he heard while I was pregnant [PM2] and during the first months of being born in France. And that's the music he'd prefer to sing. [MP3]

R: would you say that he has favourite songs by now?

G: Yes [MP4]

R: and when did he learned them?

G: when we got here [Cambridge] he was 6 months old, so we started to go to playgroups [MP4]. And there is this French work that does actually the song that I use to put him and they sing it and they dance at it. And now he knows what to do with each song, even things that I didn't teach him. Sometimes I start singing the same song in English and he'd say 'no', like 'you have to sing it in French'

C: Twinkle twinkle, Itsy-Bitsy Spider, the Crocodile song… more than five. [MP1]

R: when did you realize she had favourite songs, or that there where songs she remembers?

C: Twinkle-twinkle was some five months ago [1.5 yo], more or less when she could talk. [MP2 LA3]

R: did she la la la the songs before that?

C: could be… can’t remember. She would ask me for songs [MP3], whose lyrics she couldn’t really remember. I do remember more lately, now that she sings.

R: and before any singing and la la la, did you noticed any particular preference for a given piece of music?

C: yes, around the first year [MP4]. It was some Disney animated movies, because there was video as well as sound. [MM1]

R: I see

M: Disney movies are always a big hit. My little girl’s crazy about [Disney’s] Frozen and sings ‘Let it go, let it goooo’ while she pretends to be the princess in movie [MM2]. Now it’s become a sort of earworm of my own [everybody laughs and nods in agreement]

X: That’ll be Elsa. I guess I replay the film’s scenes in my head too [chuckles]

[Researcher debriefs participants]

Appendix H. Vocal data

Participant 1, Hugh

P1. Mother ADSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.00	5.89	472	495	182.385	957.4	715.4	3.9	1.4	2.5	56.66
2	15.25	5.341	220	226	66.448	510	377.3	5.2	0	5.2	64.94
3	15.50	5.652	223	237	80.308	538	389.4	1.9	0	1.9	51.43
4	15.75	5.694	231	231	40.642	309.9	158.1	8.5	0	8.5	59.01
5	16.00	5.077	233	239	45.777	336.1	170.5	5	5	0	64.5
6	16.25	5.396	238	254	71.933	504.3	359.1	10	7.5	2.5	68.85
8	16.75	5.502	261	282	90.988	565.6	397.2	7.1	0	7.1	60.32
9	17.00	5.324	230	248	90.04	613.5	446.9	0	0	0	80.18
10	17.25	6.982	230	238	62.815	443.4	282.2	0.8	0	0.8	58.79
11	17.50	6.459	229	247	91.116	571.2	429.2	0	0	0	60.98
12	17.75	5.062	230	233	49.959	373.5	231.5	20	0	20	65.41
13	18.00	5.535	209	206	36.277	279.4	165.1	4.9	1.2	3.7	70.29
14	18.25	6.057	200	214	61.244	451.2	315.1	6.9	1.7	5.2	65.65
15	18.50	5.661	200	210	52.073	367.5	221.3	7.5	1.5	6	71.38
16	18.75	5.2	228	240	60.045	474.9	300.9	2.1	0	2.1	80.65
18	19.25	5.286	208	221	73.991	490.4	356.1	1.5	0	1.5	61.92
19	19.50	4.303	248	241	36.059	286.3	97.7	12.5	0	12.5	88.03
20	19.75	4.733	181	180	31.662	234.5	131.5	6.5	0	6.5	59.95

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P1. Child vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.00	5.89	472	495	182.385	957.4	715.4	3.9	1.4	2.5	56.66
2	15.25	5.439	450	446	119.936	853.9	586.8	2.4	0	2.4	68.45
3	15.50	5.867	428	434	92.819	660.2	351.7	0	0	0	66.79
4	15.75	5.511	404	411	104.028	616.2	402.5	2.9	0.3	2.6	72.91
5	16.00	6.597	354	385	129.059	802.5	570.1	4.1	1.4	2.7	64.7
6	16.25	4.651	392	395	90.901	582.7	378	6.1	3	3	71.35
8	16.75	5.82	395	407	120.318	781.6	553.6	8.6	2.9	5.7	75.69
9	17.00	6.323	316	335	72.486	498.7	286.6	3.7	0	3.7	57.81
10	17.25	7.266	388	434	161.681	844	586.7	2.6	0.7	2	61.69
11	17.50	6.611	476	469	107.083	716.9	424.6	0	0	0	55.86
12	17.75	6.957	478	479	133.59	843.8	574.1	6	0	6	57.19
13	18.00	5.826	456	465	130.847	700.6	433.7	5.5	2.5	3	63.94
14	18.25	6.497	382	389	81.042	614.1	349.1	0.9	0	0.9	63.34
15	18.50	4.529	436	425	95.704	576.2	329	4.6	0	4.6	78.19
16	18.75	4.156	344	359	98.263	760.1	509.1	6.2	0	6.2	52.2
17	19.00	5.115	445	446	113.795	817.8	560	16.2	0	16.2	71.95
18	19.25	4.924	396	412	87.064	642.6	380.4	3.7	1.2	2.5	57.15
19	19.50	6.681	461	476	129.697	765.5	465.4	3	1.5	1.5	49.56
20	19.75	4.958	370	387	122.036	660.8	425.8	3.3	0	3.3	63.97

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P1. Mother IDSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	IDSp/ADSp	Utt/min
1	15.00	4.61	222	251	96.39	510.6	380.3	6.2	2.3	4.5	61.24	36.0	5.3
2	15.25	5.624	279	293	96.38	490.9	335.3	7.1	1.4	5.7	65.41	6.6	5.2
3	15.50	5.991	344	356	106.403	638.5	448.5	1	1	0	57.74	1.4	3.0
4	15.75	5.933	254	270	86.613	496.1	343.1	4.7	0.7	4	63.17	4.3	5.2
5	16.00	5.552	249	275	98.392	502	355.4	7.8	2.6	5.2	56.44	11.2	9.0
6	16.25	4.568	297	304	80.804	472.2	294.7	12.2	0	12.2	56.09	2.8	1.3
8	16.75	5.321	261	290	115.442	653.1	510.1	9.1	0	9.1	73.1	1.6	1.4
9	17.00	6.671	262	285	95.647	526.2	359.7	4.5	1.9	2.6	63.09	4.8	4.2
10	17.25	6.383	289	323	131.293	646.3	484.1	3.8	0.6	3.2	57.84	2.0	3.8
11	17.50	5.994	302	313	112.32	649.2	472.2	1	0	1	59.1	1.7	2.5
12	17.75	5.889	268	294	103.187	576.3	411.3	15.3	8.2	9.2	66.75	2.8	1.7
13	18.00	5.063	265	285	108.139	637.6	484.6	10.5	1.8	8.8	66.08	1.3	3.0
14	18.25	6.318	242	273	99.484	572.2	417.2	11	3.1	7.9	67.16	2.3	2.7
15	18.50	5.332	307	327	111.508	658.5	494.3	14.5	5.8	8.7	65.16	1.1	1.6
16	18.75	4.739	364	372	118.025	641.4	456.4	26.3	10.5	15.8	78.15	1.0	0.8
17	19.00	5.451	268	291	125.338	635.6	497.7	11.7	1.7	10	67.32	8.0	1.3
18	19.25	5.077	264	293	105.354	637.6	478.2	12.8	3.2	9.6	58.78	1.2	1.8
19	19.50	5.372	368	370	115.754	618	428.1	20.6	8.8	11.8	79.91	3.0	0.8
20	19.75	5.662	388	371	85.822	481.9	371	16.7	16.7	0	57.31	0.3	0.3

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
IDSp/ADSp = ratio between the total number of IDSp utterances and the total number of ADSp utterances
Utt/min = number of IDSp utterances per minute
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P1. Mother IDSp Enquiring vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	15.25	4.64	302	309	82.674	457.8	290.7	30	6.7	23.3	55.6
3	15.50	5.875	362	374	114.231	851.3	601.6	4.2	4.2	0	57.65
4	15.75	7.025	256	275	86.147	501.6	340.9	5.8	1.9	3.8	63.76
5	16.00	5.851	289	300	82.11	495.3	316.3	5.9	1.2	4.7	62.22
6	16.25	5.963	302	322	70.588	488.4	281.6	0	0	0	37.89
8	16.75	5.09	244	255	70.527	397.6	257.7	11.5	0	11.5	70.25
9	17.00	5.998	278	292	84.519	478.3	306.4	5.7	2.3	3.4	54.46
10	17.25	7.004	286	310	118.361	593.6	429.6	1	0	1	64.68
11	17.50	5.94	242	263	72.004	492	316.7	0	0	0	62.45
12	17.75	5.633	262	293	100.736	577.4	410.3	15.1	9.6	8.2	67.28
13	18.00	5.24	248	279	98.892	589.6	424.4	0	0	0	49.75
14	18.25	6.386	254	276	86.975	518.9	345.8	7.3	4.9	2.4	55.54
15	18.50	5.903	341	343	96.286	511.4	343.9	18.8	12.5	6.2	56.62
16	18.75	3.445	290	279	56.763	374.2	211.1	16.7	0	16.7	101.87
17	19.00	5.556	316	329	92.109	579.3	418.8	6.7	0	6.7	82.3
18	19.25	5.033	295	305	100.708	611.8	441.3	15.6	6.2	9.4	64.26
19	19.50	5.264	371	367	106.775	591.2	392.4	13.6	13.6	0	68.9
20	19.75	6.848	199	227	82.767	453.9	332.3	8.3	8.3	0	68.74

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P1. Mother IDSp Gratifying vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	15.25	3.871	228	232	36.358	272.4	109.1	0	0	0	3.871
3	15.50	4.833	294	276	83.084	416.7	270.6	0	0	0	4.833
4	15.75	5.372	188	196	39.439	294.4	156.8	3	0	3	5.372
5	16.00	4.044	176	191	70.06	476.9	378.7	2.4	0	2.4	4.044
6	16.25	3.972	285	293	73.297	441.8	261.8	30	0	30	3.972
9	17.00	5.433	186	197	44.3	282.9	146.7	0	0	0	5.433
10	17.25	5.355	234	238	78.081	538.7	408	17.6	0	17.6	5.355
11	17.50	5.447	253	309	136.146	615.5	448.5	0	0	0	5.447
13	18.00	4.034	192	194	36.759	283.4	145.6	0	0	0	4.034
14	18.25	5.008	192	215	65.568	440.7	276.6	0	0	0	5.008
15	18.50	2.121	218	220	76.999	331.3	211.9	0	0	0	2.121
16	18.75	6.092	430	427	110.93	640.2	396.9	46.2	15.4	30.8	6.092
17	19.00	5.263	192	188	40.819	274.9	156.8	0	0	0	5.263

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

Participant 2, Louisa

P2. Mother IDSp Pedagogic vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	% IDSp Ped
1	15.00	5.089	207	231	88.79	498.5	375.8	6.6	0	6.6	59.53	36.1
2	15.25	5.589	267	275	91.606	459	322.2	5.4	1.1	4.3	63.4	24.5
3	15.50	6.4	317	331	102.923	630.3	462.6	0	0	0	60.96	26.7
4	15.75	6.004	268	282	84.301	500.9	338.9	4.1	0.7	3.4	57.85	27.7
5	16.00	5.615	288	298	91.876	502.8	354.1	12.5	3.7	8.8	61.21	44.6
6	16.25	4.118	325	315	90.052	479	301.8	15.8	5.3	10.5	72.94	36.4
8	16.75	4.515	243	242	57.471	375.1	233.8	15.4	0	15.4	80.43	12.5
9	17.00	6.344	285	291	79.742	468.7	282.6	0	0	0	67.8	13.8
10	17.25	6.222	236	280	105.439	524.7	378.9	6.7	0	6.7	60.09	19.5
11	17.50	6.179	309	307	73.674	485.8	299.7	1.7	0	1.7	62.73	50.0
12	17.75	6.259	270	283	82.084	461.4	294.7	10.2	4.1	6.1	57.31	35.3
13	18.00	5.228	274	291	107.814	679.3	521.3	5.7	1.9	3.8	58.27	30.0
14	18.25	6.023	237	268	106.265	592.9	451.6	15.3	6.9	8.3	63.18	59.3
15	18.50	6.22	241	258	86.93	509.1	384	0	0	0	56.77	10.0
17	19.00	5.491	259	257	75.238	408.9	257.9	10.3	2.6	7.7	61.66	37.5
18	19.25	4.992	269	313	132.568	651.8	507.5	30.8	7.7	23.1	49.67	20.0

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother IDSp Playful vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.00	4.176	240	264	97.29	505.7	370.6	9.8	4.3	6.5	60.68
2	15.25	5.259	267	287	106.168	611.9	467.8	3.5	0.7	2.8	66.17
4	15.75	5.652	226	249	81.802	459.2	306.1	3.5	0	3.5	61.72
5	16.00	5.019	218	236	102.951	508.3	385.3	4.5	0	4.5	51.32
8	16.75	5.025	210	208	29.987	265.2	97.1	0	0	0	61.28
9	17.00	6.668	265	300	115.535	575.8	407.5	4.7	1.2	3.5	61.56
10	17.25	7.198	292	331	142.24	655.2	490.6	3.9	1	2.9	56.97
11	17.50	6.689	270	300	107.204	611.7	433.1	0	0	0	60.79
13	18.00	5.269	291	305	75.942	448.7	263.8	16.7	5.6	11.1	51.74
14	18.25	5.496	281	302	116.266	604.7	455.3	15.6	0	15.6	79.19
15	18.50	5.479	410	393	76.03	523.2	350.7	10	0	10	51.77
16	18.75	3.347	317	317	78.553	487.4	332.5	18.2	0	18.2	96.51
18	19.25	5.029	257	286	101.138	600.6	445.7	7	1.4	5.6	54.25
19	19.50	5.291	368	368	123.133	632.6	445	23.3	10	13.3	80.71
20	19.75	5.877	389	363	91.802	489.6	279.8	0	0	0	51.5

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother ADSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.25	4.851	210	225	80.03	542.6	400.1	5.6	0	5.6	52.29
2	15.35	5.024	180	199	73.605	497.1	354.7	0	0	0	96.56
3	15.50	4.428	129	135	24.576	208.1	116.1	0	0	0	61.96
4	15.75	4.262	163	163	22.197	220.8	107.6	0	0	0	54.8
5	16.00	5.405	148	154	16.097	183.3	42.6	0	0	0	55.53
7	16.50	4.923	203	205	42.841	310.6	203.6	3.4	3.4	0	64.79
8	16.75	5.467	176	171	36.817	232.8	137.1	0	0	0	59.58
9	17.00	7.467	194	199	56.088	427.7	325.5	0	0	0	60.03
10	17.25	5.211	199	201	20.945	246	81.7	4.8	0	4.8	51.06
11	17.50	5.147	210	208	26.874	263.5	91.5	0	0	0	45.31
12	17.75	4.926	186	188	51.483	399	293.4	0	0	0	66.45
13	18.00	5.389	176	173	16.795	194.2	51.4	14.3	0	14.3	60.96
14	18.25	5.593	180	183	35.045	289	157.6	0	0	0	80.08
15	18.50	5.256	180	225	111.03	405.7	275.1	0	0	0	98.14
16	18.75	5.612	196	229	111.84	553.3	423.1	0	0	0	62.98
17	19.00	5.578	202	210	76.768	557.1	427.5	3.7	3.7	0	72.93
18	19.25	5.741	196	207	75.886	518.2	384.5	5	1.7	3.3	72.1
19	19.50	4.149	174	195	87.444	470.6	375.1	0	0	0	75.23

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Child vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.25	5.89	472	495	182.385	957.4	715.4	3.9	1.4	2.5	56.66
2	15.35	5.439	450	446	119.936	853.9	586.8	2.4	0	2.4	68.45
3	15.50	5.867	428	434	92.819	660.2	351.7	0	0	0	66.79
4	15.75	5.511	404	411	104.028	616.2	402.5	2.9	0.3	2.6	72.91
5	16.00	6.597	354	385	129.059	802.5	570.1	4.1	1.4	2.7	64.7
6	16.25	4.651	392	395	90.901	582.7	378	6.1	3	3	71.35
7	16.50	5.82	395	407	120.318	781.6	553.6	8.6	2.9	5.7	75.69
8	16.75	6.323	316	335	72.486	498.7	286.6	3.7	0	3.7	57.81
9	17.00	7.266	388	434	161.681	844	586.7	2.6	0.7	2	61.69
10	17.25	6.611	476	469	107.083	716.9	424.6	0	0	0	55.86
11	17.50	6.957	478	479	133.59	843.8	574.1	6	0	6	57.19
12	17.75	5.826	456	465	130.847	700.6	433.7	5.5	2.5	3	63.94
13	18.00	6.497	382	389	81.042	614.1	349.1	0.9	0	0.9	63.34
14	18.25	4.529	436	425	95.704	576.2	329	4.6	0	4.6	78.19
15	18.50	4.156	344	359	98.263	760.1	509.1	6.2	0	6.2	52.2
16	18.75	5.115	445	446	113.795	817.8	560	16.2	0	16.2	71.95
17	19.00	4.924	396	412	87.064	642.6	380.4	3.7	1.2	2.5	57.15
18	19.25	6.681	461	476	129.697	765.5	465.4	3	1.5	1.5	49.56
19	19.50	4.958	370	387	122.036	660.8	425.8	3.3	0	3.3	63.97

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother IDSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	IDSp/ADSp	Utt/min
1	15.25	5.439	258	282	109.069	563.3	421.3	2.7	0	2.7	58.69	2.9	5.6
2	15.35	5.623	242	242	67.223	370.9	234.1	3.6	3.6	0	46.34	1.6	2.0
4	15.75	4.686	242	260	83.428	519.5	370.8	8.3	1.4	6.9	72.16	6.0	3.9
5	16.00	5.904	221	237	76.352	498.2	371.6	7.4	5.6	1.9	68.27	2.7	3.0
6	16.25	5.888	221	261	103.012	507.1	385.8	1	1	0	63.66	8.5	3.1
7	16.50	4.901	211	232	82.855	489	370.4	6.6	2.9	4	59.85	6.0	7.7
8	16.75	5.521	207	227	74.782	425	306.7	2.7	0.3	2.3	62.8	5.2	10.3
9	17.00	6.406	337	392	184.768	848.7	671.5	1.7	1.2	0.6	58.66	5.3	6.9
10	17.25	5.688	201	218	74.965	422.1	309	6.9	4.2	2.8	59.87	1.6	3.6
11	17.50	5.779	266	312	151.033	735.8	586.5	2.5	2	0.5	70.6	16.3	9.6
12	17.75	5.443	214	233	90.738	519.5	394.5	3.6	0.7	2.9	74.33	2.3	5.1
13	18.00	5.146	228	266	119.41	566	430.1	3.6	1.8	1.8	68.36	9.0	4.9
14	18.25	5.473	221	231	63.717	394	263.9	6.7	1.7	5	58.84	5.5	6.6
15	18.50	6.496	262	310	145.131	658.5	518.5	0.4	0.4	0	63	44.0	9.0
16	18.75	4.67	209	238	89.492	511.1	387.2	8.2	2.8	5.4	64.25	15.5	6.3
17	19.00	5.578	202	210	76.768	557.1	427.5	3.7	3.7	0	72.93	3.3	4.8
18	19.25	5.735	204	239	93.711	518.9	381.9	6.2	4.1	2.1	65.16	2.5	7.1
19	19.50	4.966	190	208	70.196	459.4	343.4	11.7	4.1	7.6	72.89	4.4	8.9

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
IDSp/ADSp = ratio between the total number of IDSp utterances and the total number of ADSp utterances
Utt/min = number of IDSp utterances per minute
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother IDSp Enquiring vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.25	7.003	240	252	38.535	352.2	137	33.3	33.3	0	42.86
2	15.35	4.836	312	372	229.77	890.4	753.3	0	0	0	45.37
4	15.75	6.846	217	217	45.302	310	168.9	0	0	0	35.94
5	16.00	5.808	258	267	94.209	691	522.1	0	0	0	63.79
6	16.25	6.228	350	337	99.561	580.5	398.7	5.6	5.6	0	55.87
7	16.50	5.556	243	247	56.034	457.8	291.2	3.1	0	3.1	54.24
8	16.75	5.875	238	233	51.276	341	202	0	0	0	67.42
9	17.00	7.106	257	286	94.613	532	346	0	0	0	46.52
10	17.25	5.901	222	235	71.883	439.4	292.8	18.2	18.2	0	42.01
12	17.75	5.424	194	204	55.331	310.9	175.9	14.3	7.1	7.1	82.91
13	18.00	4.967	226	242	95.51	514.1	387.3	4.2	0	4.2	76.18
14	18.25	4.735	216	217	49.795	359.5	237	9.1	9.1	0	48.76
15	18.50	6.355	249	278	113.402	617.7	446.5	0	0	0	39.06
16	18.75	6.032	214	268	124.32	564.4	407.9	0	0	0	78.8
17	19.00	5.408	228	232	30.261	316.5	129.5	4.3	4.3	0	74.2
18	19.25	6.494	248	304	146.072	725.1	569.8	5.6	5.6	0	58.96
19	19.50	5.173	212	227	57.807	377.8	230	4.5	4.5	0	69.29

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother IDSp Gratifying vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
5	16.00	4.644	144	155	26.681	191.9	76.1	0	0	0	62.09
6	16.25	4.42	220	241	71.12	453.1	296.5	0	0	0	97.04
10	17.25	5.828	152	156	27.334	202.9	89.8	0	0	0	69.31
11	17.50	7.129	322	354	160.448	871.4	697.3	0	0	0	65.02
12	17.75	4.778	222	221	41.879	283.4	125	0	0	0	123.42
14	18.25	5.457	204	213	51.356	330.9	191.6	31.6	5.3	26.3	75.09
15	18.50	5.779	264	289	128.417	692.5	547.4	0	0	0	46.42
17	19.00	4.699	169	192	70.861	412	289.9	20	0	20	97.91
18	19.25	4.37	171	171	21.414	218	91.8	18.2	18.2	0	91.49
19	19.50	5.087	185	189	59.535	363.5	263.5	28	4	24	78.46

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother IDSp Pedagogic vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	% IDSp Ped
1	15.25	6.317	324	394	197.879	902.9	708.4	2	2	0	68.2	17.4
2	15.35	7.153	286	358	187.706	789.3	642.2	0	0	0	52.34	45.5
4	15.75	4.869	292	292	44.729	368	150.2	0	0	0	147.37	5.6
5	16.00	5.491	194	196	49.638	387.2	267.1	5.9	5.9	0	73.86	25.0
6	16.25	5.834	185	214	76.608	415.6	295.2	0	0	0	73.31	23.5
7	16.50	6.033	261	284	88.591	500.7	338.4	6.7	2.2	4.4	67.26	22.2
8	16.75	5.269	174	191	53.533	334.2	224.4	8.6	5.7	2.9	86.6	12.9
9	17.00	6.051	223	229	58.888	481.3	356.3	0	0	0	67.66	31.3
10	17.25	5.645	233	238	82.362	538.3	418.4	2.5	2.5	0	50.04	50.0
11	17.50	6.931	276	368	182.131	763.1	588.5	4.3	4.3	0	69.63	22.4
12	17.75	5.21	224	251	111.541	627.7	500.4	2.3	0	2.3	79.36	35.7
13	18.00	5.316	220	276	127.364	591.9	451.4	2.7	1.4	1.4	66.61	59.3
14	18.25	5.018	216	218	49.892	368.6	248.4	10.1	7.6	3.8	57.48	60.6
15	18.50	5.815	220	238	94.9	533.9	398.8	0	0	0	70.25	29.5
16	18.75	4.319	222	238	85.155	452.6	347.3	13	5.6	7.4	71	41.9
17	19.00	6.041	191	206	68.327	485	360.8	0	0	0	67.23	30.4
18	19.25	5.291	192	214	82.757	514.8	389.4	7	5	2	71.19	36.8
19	19.50	4.938	186	207	70.931	444.1	324.5	4.9	2.9	2	67.84	54.5

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P2. Mother IDSp Playful vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	15.25	5.539	232	276	125.182	592.6	449.6	1.1	0	1.1	60.9
2	15.35	4.317	232	219	51.045	280.5	160.7	0	0	0	101.6
4	15.75	4.444	262	297	110.864	548.9	399.1	12.5	0	12.5	75.06
5	16.00	5.076	246	260	122.475	532	408.4	11.1	0	11.1	64.27
6	16.25	6.112	217	266	101.827	471	325.8	0	0	0	64.77
7	16.50	4.14	272	321	99.193	502.1	277.6	0	0	0	86.24
8	16.75	6.391	274	286	88.062	452.5	294.7	0	0	0	66.86
9	17.00	6.645	436	482	189.741	887.9	677.9	2	1	1	62.06
10	17.25	3.588	152	162	34.368	223.9	116.5	16.7	0	16.7	101.6
11	17.50	4.773	266	284	105.11	680.3	493.6	0	0	0	73.8
12	17.75	5.461	204	232	96.334	535.6	407.8	4.7	0	4.7	71.54
13	18.00	4.678	168	172	29.486	246.4	120.2	0	0	0	33.06
14	18.25	4.091	214	249	89.097	410.9	274.2	0	0	0	59.74
15	18.50	7.224	382	406	159.227	699.6	511.4	0	0	0	65.28
18	19.25	5.6	214	254	99.36	521.6	386.6	4.2	1.7	2.5	58.84
19	19.50	3.702	174	200	85.443	476.9	383.6	4.8	0	4.8	88.64

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

Participant 3, Daisy

P3. Mother ADSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
4	17.0	5.661	230	218	37.606	273.4	123.4	0	0	0	46.12
5	17.3	6.074	234	233	6.743	243.5	20.6	0	0	0	40.52
7	17.8	7.432	194	194	26.77	251.9	100.8	0	0	0	29.61
8	18.0	4.75	365	413	122.607	730.6	439.6	5.3	5.3	0	63.26
9	18.25	2.827	194	196	9.883	215.5	30.6	0	0	0	56.6
10	18.5	4.935	96	96	2.517	98.8	8.5	0	0	0	175.38
13	19.3	5.67	180	216	105.486	490.1	323.3	0	0	0	49.11
14	19.5	3.367	160	160	6.137	170.1	15.5	0	0	0	149.21
15	19.75	5.536	375	413	139.426	867.6	595.6	0	0	0	66.59

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P3. Child vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	16.5	5.776	338	393	143.11	917	662.7	3.8	1.9	1.9	75.64
3	16.75	7.688	346	396	132.108	768.8	530.6	0	0	0	60.72
4	17.0	6.622	317	350	107.559	751	541.9	0.7	0	0.7	66.01
5	17.3	6.616	396	422	142.949	751.9	528.9	4.8	2.4	4.8	60.41
8	18.0	6.383	454	479	142.141	786.9	506	4.7	0	4.7	58.66
9	18.25	7.034	582	569	197.039	963.1	721.1	3.7	1.2	2.5	63.46
10	18.5	6.297	327	374	117.209	781.2	523.4	5.4	2.7	5.4	56.23
11	18.8	6.109	354	395	138.587	888.3	690.3	2.5	0.8	1.7	65.81
12	19	5.501	368	381	89.797	649.3	386	2.6	0.9	1.7	67.46
13	19.3	4.869	334	372	127.679	808	570.5	6	0	6	70.59
15	19.75	5.145	268	293	75.053	559.7	346.8	0	0	0	72.48
16	20.0	6.494	408	438	157.987	961.5	708.9	1.1	0	1.1	60.99
17	20.3	5.402	388	396	80.816	643	373.9	4.5	3.4	1.1	69.86
18	20.5	5.887	462	470	113.582	749.2	457.2	0	0	0	63.33
19	20.8	5.034	354	369	50.64	473.5	216.5	4.5	0	4.5	101.04
20	21.0	4.859	334	352	99.798	703.8	476.6	2.3	0.5	1.8	64.74

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P3. Mother IDSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	IDSp/ADSp	Utt/min
2	16.5	6.307	290	287	105.935	505.9	357.9	6.4	5.3	1.1	69.28	32.0	8.5
3	16.75	6.331	278	287	88.281	486.3	328.3	5.7	3.6	2	63.7	12.7	6.6
4	17.0	6.325	292	310	113.905	686	515	0.4	0.4	0	63.92	42.0	7.6
5	17.3	6.412	310	319	102.668	665.1	508.1	3.5	2.9	0.6	64.97	35.0	7.9
7	17.8	6.051	270	288	98.116	617.9	452.8	7.6	3.3	4.3	55.42	18.0	4.0
8	18.0	5.552	266	278	82.753	533.4	377.4	11.7	6.9	5.5	69.84	33.0	7.9
9	18.25	5.914	274	286	99.988	632.4	477.3	7.4	5.1	2.8	63.83	41.0	7.7
10	18.5	5.206	283	284	99.284	617.7	474.3	16.9	11.9	5.9	69.3	26.0	5.1
11	18.8	5.717	236	249	78.255	427.8	283.4	9.6	4.1	5.9	63.88	34.0	6.4
12	19	5.458	275	285	96.471	496	342.5	9.4	4.3	5.4	60.57	53.0	10.4
13	19.3	4.688	248	269	82.921	466.8	306.8	8.9	4.8	4.2	67.61	67.0	13.4
14	19.5	5.521	270	282	76.774	479.3	302.2	8.4	3.7	4.7	66.69	40.0	6.9
15	19.75	5.821	274	283	80.388	462.7	299.7	15.4	12.5	2.9	58.99	23.0	4.5
16	20.0	5.773	300	314	97.176	564.5	391	4.8	2.8	2	59.75	44.0	7.9
17	20.3	5.342	259	266	54.806	389.8	216.2	6.3	2.5	3.8	72.96	30.0	5.6
18	20.5	5.109	279	294	83.977	475.6	303.6	6.8	1.7	5.1	57.27	26.0	5.0
19	20.8	5.579	247	263	93.112	659.7	490.7	10.3	3.4	6.9	61.56	28.0	5.6
20	21.0	4.227	218	229	59.044	408.1	257.2	9.1	2.7	7	63.02	13.7	8.1

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
IDSp/ADSp = ratio between the total number of IDSp utterances and the total number of ADSp utterances
Utt/min = number of IDSp utterances per minute
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P3. Mother IDSp Enquiring vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	16.5	6.535	330	330	68.646	502.3	296.8	0	0	0	54.61
3	16.75	6.734	324	326	75.059	479.7	265.7	3.7	3.7	0	64.36
4	17.0	6.843	318	328	84.703	544.4	345.5	0	0	0	64.01
5	17.3	6.978	308	336	92.472	672.2	417	2.5	2.5	0	57.81
7	17.8	6.395	272	299	103.187	636.6	456.4	6.8	4.5	2.3	48.85
8	18.0	5.482	250	263	57.724	355.6	182	5.6	0	5.6	41.42
9	18.25	6.679	302	312	98.641	637.2	481.3	2.3	2.3	0	56.15
10	18.5	5.67	308	294	49.371	354	175.8	0	0	0	71.21
11	18.8	6.08	268	281	87.205	628.8	465.9	12.1	5.5	7.7	62.82
12	19	6.544	335	347	86.759	565.7	396	5.9	5.9	0	52.67
13	19.3	5.203	346	344	82.045	501.6	330.4	0	0	0	68.57
14	19.5	4.762	318	320	74.591	493	304.8	10.3	7.7	2.6	57.12
15	19.75	5.637	276	292	75.37	467.4	306.4	18.4	15.8	2.6	45.78
16	20.0	6.061	326	330	103.281	628.6	458.6	4.6	3.4	1.1	46.2
17	20.3	6.327	282	284	48.108	382.8	194.9	1.7	0	1.7	64.36
18	20.5	4.935	291	293	63.69	449.1	255.4	5.7	1.9	3.8	66.32
19	20.8	4.935	291	293	63.69	449.1	255.4	5.7	1.9	3.8	66.32
20	21.0	5.196	286	312	86.326	452.8	297.3	7.7	7.7	0	74.82

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P3. Mother IDSp Gratifying vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	16.5	3.736	182	192	34.551	279.8	123.2	20	0	20	96.1
3	16.75	6.104	202	215	49.986	355.5	203.5	16	8	8	59.93
4	17.0	6.011	291	280	47.617	380.7	205.6	0	0	0	59.39
5	17.3	4.104	200	231	145.303	558.3	451	14.3	0	14.3	72.21
8	18.0	2.842	260	280	138.144	477	317.5	0	0	0	113.04
9	18.25	4.797	208	246	104.584	534.7	384.2	19.4	6.5	12.9	71.17
10	18.5	5.082	264	293	133.677	650.8	507.7	25.9	22.2	3.7	67.39
11	18.8	5.996	218	221	54.64	342.4	193.5	17.6	8.8	8.8	66.74
12	19	5.766	264	273	83.067	447	289	13.5	3.8	9.6	60.2
13	19.3	5.074	234	249	75.331	473	313	16.7	7.8	8.9	69.55
15	19.75	4.901	314	325	85.043	480.8	275.2	44.4	33.3	11.1	80.3
16	20.0	6.008	270	280	96.942	664.5	522.2	0	0	0	68.25
17	20.3	4.449	205	218	51.696	389.4	222.1	13.6	4.5	9.1	66.07
20	21.0	3.609	194	210	47.334	341.1	179.4	25	12.5	25	79.21

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P3. Mother IDSp Pedagogic vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	% IDSp Ped
2	16.5	6.348	279	282	103.029	497.4	349.5	8.1	7.5	0.6	68.79	81.3
3	16.75	6.51	288	300	96.566	568.7	406.7	6.9	5.8	1.1	67.23	65.8
4	17.0	6.309	296	312	115.902	689.6	516.9	0.6	0.6	0	66.97	69.0
5	17.3	6.246	308	303	82.765	482.9	326.9	4.4	4.4	0	63.92	68.6
7	17.8	5.473	278	279	79.893	519.7	359.9	18.9	8.1	10.8	63.5	50.0
8	18.0	5.419	264	276	90.464	635.1	491.7	11	6.3	5.5	69.88	69.7
9	18.25	6.091	275	288	98.294	637.4	481.4	7.5	5.7	2.5	63.95	65.9
10	18.5	5.14	262	268	82.903	526.7	383.2	8.9	4.4	6.7	71.16	42.3
11	18.8	6.067	261	268	97.054	548.9	404.5	11.1	5.1	6	58.49	52.9
12	19	5.597	276	281	92.129	475.2	328.5	8.1	5.4	3.4	58.74	54.7
13	19.3	4.948	254	275	85.973	497.8	338.8	10.2	6.2	3.9	69.33	55.2
14	19.5	5.657	261	280	86.931	506.8	334.9	8.3	4.2	4.2	64.34	47.5
15	19.75	5.922	272	285	83.913	478.1	316.9	17.9	12.6	5.3	60.02	87.0
16	20.0	5.756	312	320	88.306	501.4	327.4	6.2	3.4	2.8	58.7	50.0
17	20.3	5.215	261	271	54.927	398	227.4	7.3	3.6	3.6	70.72	70.0
18	20.5	5.248	260	279	85.989	504.7	338.1	5.8	1.2	4.7	63.79	61.5
19	20.8	5.444	268	285	84.841	509.4	336.6	4.9	1.2	3.7	65.82	57.1
20	21.0	4.413	221	236	62.755	413.7	261.2	11.6	4.7	7	66.38	34.1

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P3. Mother IDSp Playful vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	16.5	9.091	206	206	48.666	254.9	96.8	0	0	0	81.82
3	16.75	6.897	283	283	83.609	452.4	291	0	0	0	54.11
4	17.0	6.571	334	313	108.776	542.1	374.8	2	0	2	66.63
5	17.3	8.187	306	310	99.343	508.7	347.1	14.3	14.3	0	74.38
7	17.8	2.891	337	348	98.355	479.3	224.9	0	0	0	114.75
8	18.0	4.887	280	301	81.235	489.2	293.4	15.8	5.3	10.5	79.23
9	18.25	6.751	312	315	72.188	483.2	284.2	6.7	0	6.7	51.31
10	18.5	4.117	213	253	101.582	479.9	340.8	31.2	25	6.2	63.04
11	18.8	5.459	220	231	57.926	351.3	205.9	14.3	2.9	14.3	71.73
12	19	5.633	262	303	133.498	593.6	439.2	12.7	5.5	7.3	60.38
13	19.3	3.996	262	309	109.134	651.3	469.8	22.9	14.3	11.4	74.38
14	19.5	5.157	279	278	61.739	391.1	211.7	17.1	2.9	14.3	54.8
16	20.0	6.606	380	368	91.329	511.5	317.2	0	0	0	64.59
17	20.3	4.669	258	251	37.838	319	150.6	9.1	0	9.1	86.71
18	20.5	3.721	174	188	31.407	238.7	86.8	50	0	50	49.7
19	20.8	3.721	174	188	31.407	238.7	86.8	50	0	50	49.7
20	21.0	2.321	179	177	20.849	198.8	49.8	0	0	0	13.16

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

Participant 4, Tommy

P4. Mother ADSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	24	5.923	267	277	69.063	492.6	339.6	3	0	3	67.36
3	24.3	6.89	240	247	73.94	508.2	379.9	0	0	0	42.31
4	24.5	5.878	264	264	71.777	553.9	384.8	4.3	0	4.3	73.22
5	24.75	5.18	231	232	66.67	342	229.4	5	0	5	73.08
6	25.0	4.525	159	166	51.3	229.1	122.3	0	0	0	89.14
8	25.5	5.305	225	224	56.477	356.9	234.5	0	0	0	56.28
13	26.8	6.232	253	248	34.394	297.8	172.9	7.1	0	7.1	68.35
15	27.3	6.51	222	234	69.227	451	339.7	0	0	0	72.83
16	27.5	5.54	215	212	52.091	295	193	0	0	0	50.2
17	27.75	5.669	230	236	32.92	331.8	150.4	0	0	0	65.63
19	28.3	5.894	245	247	77.525	609.3	482.6	1.3	0	1.3	64.78

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations

P4. Child vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	24	6.067	369	409	137.138	852.1	645.4	3.7	0	3.7	78.78
3	24.3	4.888	330	377	144.11	792.6	601.6	0	0	0	58.93
4	24.5	4.078	346	425	165.721	820.9	577.3	0	0	0	56.67
5	24.75	5.123	352	371	119.177	737.7	557.2	2.4	1.2	1.2	55.71
6	25.0	6.135	372	430	140.227	733.8	473.5	5	0	5	53.22
7	25.3	5.683	372	415	116.37	746.5	463.7	0	0	0	67.44
8	25.5	4.95	292	325	69.795	469.6	236	0	0	0	50.75
9	25.8	6.042	365	369	67.258	475.5	207.5	0	0	0	63.52
13	26.8	4.735	334	359	101.508	663.8	438.6	4.7	1.6	3.1	67.18
15	27.3	5.738	284	322	121.347	750.4	523.3	3.6	0	3.6	67.89
16	27.5	5.167	340	362	104.006	606.4	383.8	1.2	0	1.2	67.7

17	27.75	5.972	472	545	195.837	937.6	696.6	1.6	0	1.6	69.98

P4. Mother IDSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	IDSp/ADSp	Utt/min
2	24	5.637	292	309	95.042	544.9	377.8	11	3.9	7.1	58.27	9.7	10.7
3	24.3	5.106	275	283	77.379	463.4	307.8	4.7	2.2	2.5	55.49	11.6	13.8
4	24.5	5.206	251	270	92.615	535.2	399.1	4.8	1.6	3.2	62.8	10.3	15.8
5	24.75	5.212	279	299	99.832	596.6	437.4	13.6	7.3	6.2	61.46	17.3	19.9
6	25.0	5.094	255	271	85.303	517.8	359.4	9.5	1.5	8	52.38	30.5	23.0
7	25.3	5.313	272	290	95.4	616	460.9	6.7	3.3	3.3	55.2	23.0	15.2
8	25.5	5.211	268	290	95.908	587	434	8.5	3.3	5.2	66.32	24.3	20.4
9	25.8	5.304	298	319	98.9	614.8	435.5	10.1	3.8	6.7	60.19	46.0	15.2
13	26.8	5.43	274	308	114.365	649.3	470.6	6.4	2.1	4.4	58.86	16.6	14.3
15	27.3	5.288	244	260	72.147	478.9	325.9	5	2.3	2.7	64.14	5.5	7.9
16	27.5	5.595	253	275	83.092	513.7	372.7	3.4	0.7	2.7	62.32	8.2	14.9
17	27.75	5.239	291	307	94.381	550.3	394.4	7.9	3.6	5.1	64.28	22.3	13.9
19	28.3	5.593	286	308	103.932	653.8	484.4	6.5	3.2	3.5	61.55	6.5	8.2

IDSp/ADSp = ratio between the total number of IDSp utterances and the total number of ADSp utterances
Utt/min = number of IDSp utterances per minute

P4. Mother IDSp Enquiring vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	24	5.669	310	326	104.705	564	386.9	11.4	4.8	6.6	55.12
3	24.3	5.665	263	280	77.461	476.7	308.1	3.6	2.7	0.9	56.67
4	24.5	4.879	316	320	85.42	516.3	355.6	3.6	1.8	1.8	70.22
5	24.75	5.252	288	307	99.848	624.3	463.2	14.2	7.1	7.1	65.11
6	25.0	5.31	255	268	72.73	464	301.5	5.8	1.9	3.9	48.56
7	25.3	5.048	286	305	88.198	608.4	403.1	5.1	1.7	3.4	48.56
8	25.5	5.351	276	292	87.343	556.1	396.2	7.9	0.7	7.2	62.55
9	25.8	5.422	297	319	95.908	600.5	418.3	9.6	3	6.6	57.88
13	26.8	5.152	264	278	82.578	603.3	432.1	4.5	0	4.5	60.89
15	27.3	5.235	247	262	75.379	495.8	360.5	3.8	1.9	1.9	57.56

16	27.5	5.556	248	257	68.506	505.1	363.8	5.7	1.9	3.8	59.9
17	27.75	5.31	284	297	87.482	550	398.3	8.1	4.3	5.4	65.28
19	28.3	5.635	282	300	94.777	582.6	416.4	6.8	3.8	3	57.62

P4. Mother IDSp Gratifying vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	24	4.459	218	242	75.683	238.4	385.3	16.7	0	16.7	58.2
3	24.3	4.28	250	308	118.018	368.4	540.4	9.1	0	9.1	70.6
4	24.5	3.346	208	191	72.823	187.5	288.9	14.3	0	14.3	67
5	24.75	4.613	257	277	106.479	447.1	564.7	12.5	12.5	0	59.2
6	25.0	4.671	226	262	84.551	247.2	429.9	16.7	0	16.7	25.52
7	25.3	5.524	192	234	89.117	345.2	474.1	8.3	8.3	0	59.7
8	25.5	4.815	214	268	133.715	474.8	594.7	9.1	9.1	0	50.87
13	26.8	5.996	262	273	73.167	332.5	465.6	4.8	0	4.8	41.9
15	27.3	3.957	182	172	75.403	221.5	298.6	6.2	0	6.2	79.4
16	27.5	6.284	246	255	70.922	300.6	457.9	7.7	0	7.7	30.12
17	27.75	6.349	278	363	181.891	448.5	645.5	25	25	0	86.99
19	28.3	5.249	356	335	74.31	197.6	396.1	50	50	0	111.11

P4. Mother IDSp Pedagogic vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	% IDSp Ped
2	24	5.561	294	314	99.815	563.8	390.7	11.3	4.6	6.7	56.62	65.5
3	24.3	5.124	276	283	76.827	442	283.7	5.1	3.2	1.9	48.21	55.2
4	24.5	5.655	282	293	73.61	461.7	305.7	2.8	2.8	0	67.13	33.9
5	24.75	5.161	280	303	108.39	650.9	491.9	15.1	7.9	7.2	63.82	84.3
6	25.0	5.001	252	271	85.105	521.7	368.6	11.9	2.8	9.1	53.02	73.8
7	25.3	5.215	280	300	95.612	596.9	410.2	6.2	2.1	4.2	52.26	78.3
8	25.5	5.282	268	289	92.079	587.2	427.2	7.4	1.4	6	68.72	61.6
9	25.8	5.327	296	318	101.917	629.2	454.2	8.9	3.7	6.3	61.52	89.1
13	26.8	5.346	265	292	99.751	614.3	451.8	6.1	1.3	4.8	63.43	72.3
15	27.3	5.139	248	267	83.562	501.4	363.7	2.8	1.7	1.1	62.77	63.6
16	27.5	5.418	246	258	70.029	474.7	338.7	3.2	0.7	2.5	66.14	64.9
17	27.75	5.3	287	305	96.286	553.3	397.4	6.9	4.1	4.6	69.92	52.2

19	28.3	5.257	276	296	95.334	591.5	415.4	8.1	4.4	4	62.42	63.1

% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)

P4. Mother IDSp Playful vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
2	24	5.266	292	317	99.791	548.6	369.3	14.3	0	14.3	48.43
4	24.5	5.771	294	307	58.724	436	223.4	0	0	0	33.52
5	24.75	5.815	289	312	92.256	520.4	347.1	5.6	1.9	3.7	53.14
6	25.0	5.303	261	282	95.286	623.5	453.2	10.3	0	10.3	56.26
7	25.3	5.724	246	284	102.412	549.8	367.8	23.5	11.8	11.8	41.13
8	25.5	4.949	277	321	128.945	724	544.6	9.6	4.1	5.5	61.97
9	25.8	5.239	242	271	98.714	498	364.3	10	10	0	82.16
13	26.8	5.873	296	346	150.837	754.1	583.1	9.6	2.9	6.6	60.18
15	27.3	5.451	240	254	60.342	433.3	266.3	7.4	4.9	2.5	49.38
16	27.5	5.741	288	308	93.716	539.6	358.3	4.1	0.7	3.4	59.49
17	27.75	5.486	306	326	109.07	639.7	454.3	8.2	1.8	7.3	64.03
19	28.3	5.193	327	350	109.135	648.2	463.9	6	4	3	63.26

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

Participant 5, Patricia

P5. Mother ADSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	23.5	6.041	215	222	31.416	286.2	116.3	0	0	0	62.2
3	23.75	5.931	201	205	23.886	261.7	98.2	0	0	0	52.16
4	24.0	205.7	209	23.352	267.3	110.6	0	0	0	60.2	5.532
6	24.3	6.604	208	216	25.65	279.1	110	0	0	0	54.1
8	24.5	5.731	208.9	212	31.23	269.1	104.2	0	0	0	60.3
10	24.8	6.669	210.7	217	32.423	280.5	103.4	0	0	0	54.2

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P5. Child vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	23.5	6.151	434	427	53.274	530.8	217.5	2.6	1.3	1.3	51.71
3	23.75	4.948	339	354	77.471	609	352.7	10.5	0	10.5	87.42
4	24.0	8.41	371	394	68.5	583.5	273.6	8.1	0.7	4.7	53.01
6	24.3	8.3	390	392	69.121	590.3	300.4	9.3	1	8	76.2
8	24.5	6.1	384	390	61.212	550.11	270.22	5.4	1	4	54.21
10	24.8	4.415	295	387	62.1	556.7	296.8	5.9	1	7.3	81.79

P5. Mother IDSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	IDSp/ADSp	Utt/min
1	23.5	5.67	277	295	77.272	493.1	310.1	3.2	0.4	2.9	56.52	7.9	21.2
3	23.75	5.778	264	273	72.09	475.9	318.9	5	2.1	2.9	61.05	7.9	21.2
4	24.0	5.61	265.7	272.2	74.7	479.6	308.3	3.9	1	1.8	57	6.6	20.8
6	24.3	5.72	273	287	76.4	490.2	312	4	1.5	3.5	60	6.5	17.2
8	24.5	5.7	269	281	74.21	480.3	317	3.6	0,5	2.4	58.4	13.8	19.3
10	24.8	5.82	275.9	295.8	75.3	490.2	320.7	4.5	1.6	4	61.2	14.8	15.9

IDSp/ADSp = ratio between the total number of IDSp utterances and the total number of ADSp utterances
Utt/min = number of IDSp utterances per minute

P5. Mother IDSp Enquiring vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	23.5	5.92	274	287	67.881	473.4	292.8	3.8	0.5	3.3	52.94
3	23.75	5.971	268	273	69.77	476.4	324.1	8.7	1.7	7	60.08
4	24.0	5.91	268.6	286.2	68.0	472.9	292.3	8.3	0.6	6.6	53.4
6	24.3	5.9	271.3	283	68.5	477.3	320.2	7.1	1	5.2	59.09
8	24.5	5.95	270.8	277	70	473.1	296.4	4.9	1	4.7	55
10	24.8	5.9	273.5	273.8	70.1	477.2	324.4	3.9	1.5	3.5	60.2

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P5. Mother IDSp Gratifying vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	23.5	6.343	286	307	93.677	494.4	308.4	0	0	49.43	308.4
3	23.75	6.531	294	326	126.78	622.4	418.6	0	0	63.41	418.6
4	24.0	5.9	286.3	324.7	96.5	610.6	409.1	0.0	0.0	62.2	318.8
6	24.3	5.683	292	319	112	560.2	366.7	0	0	58	366
8	24.5	6.413	288	315	108	556.1	362.8	0	0	54	362.1
10		6.6	293.7	308.8	123.7	506.0	319.1	0.0	0.0	50.3	408.8

P5. Mother IDSp Pedagogic vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	% IDSp Ped
1	23.5	5.833	272	284	67.003	466.5	286	4	0.4	3.5	56.43	71.4
3	23.75	5.869	260	265	62.001	422.7	265.7	6	2.8	3.2	62.14	90.9
4	24.0	3.7	260.5	265.6	62.5	426.8	267.4	4.2	0.0	3.1	55.2	89.4
6	24.3	9.4	269	269	66	449.4	280.3	5	3	3.6	61.14	84.5
8	24.5	3.5	263	281	64	441.2	273	5	0	3.1	55.1	75.5
10	24.8	8.6	271.5	283.9	67.0	463.1	285.1	5.8	3.1	3.6	62.2	71.8

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P5. Mother IDSp Playful vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	23.5	5.176	342	354	117.983	676.6	465.5	5.9	5.9	0	57.23
3	23.75	5.529	248	266	79.594	440	290	2	2	0	64.03
4	24.0	5.2	333.5	273.9	83.0	654.9	449.7	2.4	5.6	0.0	57.6
6	24.3	5.53	299	314	103.123	562.3	380.6	3.8	4	0	63.11
8	24.5	5.22	291	306	95.341	554.1	376.1	4.2	4	0	59.12
10	24.8	5.6	256.5	346.1	115.0	461.6	306.4	5.6	2.4	0.0	64.1

Participant 6, Gwendolyn

P6. Mother ADSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	21.75	5.106	214	211	31.639	269.9	129.7	3.3	2.2	2.2	58.28
2	22.0	6.218	230	235	64.045	406.2	287.1	0	0	0	46.38
4	22.5	6.207	206	208	19.586	258.9	85.1	0	0	0	34.09
6	23	6.034	225	229	37.944	364.2	178.8	0	0	0	69.17
8	23.5	6.89	220	221	41.215	328.2	170.2	1	0	1	54.4
10	24	4.9	216	219	35.11	320.1	170.1	0	1.2	0	48.67

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P6. Child vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	21.75	5.049	378	378	70.288	528.4	298.3	5.6	1.9	3.8	63.52
2	22.0	5.273	398	398	73.773	607.1	342.1	9.8	0	9.8	67.68
4	22.5	6.973	418	423	89.008	701.7	447.6	2.9	0	2.9	55.58
6	23	6.237	419	428	114.139	703.5	473.8	5.6	2.1	3.5	58.95
8	23.5	8.47	408	410	90.023	646.3	393.3	4.1	1	7.21	64.16
10	24	2.341	398	404	84.042	624.2	387.1	8.2	1	3.4	58.018

P6. Mother IDSp vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	IDSp/ADSp	Utt/min
1	21.75	6.297	237	241	49.062	381.9	208.9	2.8	0.5	2.3	57.06	2.0	7.8
2	22.0	5.258	221	227	64.422	441	316.6	6.6	3.3	3.3	63.51	8.5	6.0
4	22.5	8.267	346	348	91.662	556.6	369.8	6.7	0	6.7	48.86	15.5	5.5
6	23	6.017	260	299	106.132	587.5	416.5	6.5	3.8	2.7	58.92	13.0	8.7
8	23.5	4.013	269	282	80.0123	497.3	331.5	6.3	2	4	55.23	9.8	7.9
10	24	8.111	263	276	76.301	487.1	324.6	6.7	2.2	5.2	59.012	11.7	5.8

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
IDSp/ADSp = ratio between the total number of IDSp utterances and the total number of ADSp utterances
Utt/min = number of IDSp utterances per minute
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P6. Mother IDSp Enquiring vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	21.75	5.581	237	249	59.437	386.7	234	8.8	3.5	5.3	53.28
4	22.5	6.985	278	301	83.349	512.5	329	3.7	0	3.7	52.55

P6. Mother IDSp Gratifying vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	21.75	5.659	274	304	76.065	399.4	172.3	33.3	33.3	3.7	172.3
4	22.5	4.584	264	267	34.257	345.6	124.7	0	0	58.24	124.7
6	23	5.882	274	346	162.273	779.9	590.4	11.8	0	62.39	590.4
8	23.5	6.1	273	304	94.432	511.8	293.7	20.1	18	43.24	346.5
10	24	4.2	269	308	88.152	505.2	299.3	10.9	4	39.78	245.5

P6. Mother IDSp Pedagogic vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_	% IDSp Ped
1	21.75	6.216	235	248	63.354	394.5	233.4	3.5	0.9	2.6	51.08	47.5
2	22.0	5.612	219	223	57.237	385.9	263.7	7.5	3.8	3.8	55.97	76.5
4	22.5	6.133	270	288	75.887	486.4	314	9.1	1.5	7.6	50.06	48.4
6	23	5.934	253	282	93.924	536.5	366	9.5	6.5	3	53.5	71.2
8	23.5	4.133	248	246	76.324	452.3	306	6.8	2.2	3.5	54.42	65.6
10	24	8.254	240	264	70.102	450.1	282	8.2	4.1	6.8	51.89	55.6

f0 StDev = f0 Standard deviation in Hertz
f0 Max = f0 maximum in Hertz
Gliss = f0 Glissandi
nPVI_= normalised Pairwise Variability Index (nPVI) of syllable nuclei durations
% IDSp Ped = percentage of Pedagogic IDSp utterances out of the total IDSp utterances (regardless of pragmatic function)
For further information, see section 2.3.3.1 of the third study, Routine audio recording Pre-processing and coding

P6. Mother IDSp Playful vocal parameters

Week	Age (m)	Speech Rate	f0 Median (Hz)	f0 Mean (Hz)	f0 StDev (Hz)	f0 Max (Hz)	Range (Hz)	Gliss	Rises	Falls	nPVI_
1	21.75	5.655	246	246	5.56	255.1	16.8	0	0	0	10
2	22.0	5.634	347	382	87.736	582.3	269	0	0	0	82.35
4	22.5	7.067	362	357	96.349	577.1	388.1	10	0	10	50.31
6	23	6.044	327	338	121.389	605	434	9.8	7.3	2.4	62.5
8	23.5	5.431	367	362	106.621	461.2	308	0	3	0	59.1
10	24	6.513	275	300	50.453	549.4	246	0	1	0	43.3

Appendix I. Musical Stimuli

Stimulus Category	Song name	Author	URL
Familiar pop song (Fam pop)	Every little thing	The Police	https://www.youtube.com/watch?v=sxkMFgFjC8E&frags=pl%2Cwn
Familiar Nursery Rhyme (Fam Nurs)	Twinkle, twinkle little star	traditional British	https://soundcloud.com/mothergooseclub/twinkle-twinkle-little-star
Unfamiliar pop songs (Unfam Pop)	Like a stone	Audioslave	https://www.youtube.com/watch?v=C6uqTWHzHvo&frags=pl%2Cwn
	Happy	Pharrell Williams	https://www.youtube.com/watch?v=TIC9wgt4X_E&frags=pl%2Cwn
	Mansard roof	Vampire Weekend	https://www.youtube.com/watch?v=f8LWcfa8No0&frags=pl%2Cwn
	Hey ya	Outkast	https://www.youtube.com/watch?v=-7XnDlYY9qw&frags=pl%2Cwn
	Don't stop til' you get enough	Michael Jackson	https://www.youtube.com/watch?v=ZorRGrDiMsA&frags=pl%2Cwn
	A pedir su mano	Juan Luis Guerra y los 440	https://www.youtube.com/watch?v=_xzIJtG9XXA&frags=pl%2Cwn
	All the single ladies (Put a ring on it)	Beyonce	https://www.youtube.com/watch?v=4z-bOdAdias&frags=pl%2Cwn
	Wanabee	Spice Girls	https://www.youtube.com/watch?v=U6GipCeuH6s
	Back for good	Take that	https://www.youtube.com/watch?v=eBIzCF-8xSg&frags=pl%2Cwn
	Surfin' USA	Beach boys	https://www.youtube.com/watch?v=EDb303T-B1w&frags=pl%2Cwn
	Hey Mickey	Toni Basil	https://youtu.be/WFrmapGCuhs
	Shape of you	Ed Sheeran	https://www.youtube.com/watch?v=xTvyyoF_LZY&frags=pl%2Cwn
	Pon De Replay	Rihanna	https://www.youtube.com/watch?v=PnhYHGimYkQ&frags=pl%2Cwn
	Girls just want to have fun	Cindy Lauper	https://www.youtube.com/watch?v=cMWWBigvbkY&frags=pl%2Cwn
	Lion sleeps tonight	The Tokens	https://www.youtube.com/watch?v=OQlByoPdG6c
	Gettin' Jiggy Wit It	Will Smith	https://www.youtube.com/watch?v=5VPm2yckO3k
	Play that funky music	Wild Cherry	https://www.youtube.com/watch?v=_pHT9yYFdZg
	One more time	Daft Punk	https://www.youtube.com/watch?v=A2VpR8HahKc
	Beautiful stranger	Madonna	https://www.youtube.com/watch?v=BoDVRRgyxtw
Unfamiliar nursery rhymes and cildren's songs (Unfam Nurs)	Navegando [Sailing]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-navegando-1
	Chinita Margarita [Margarita the ladybug]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-la-chinita-margarita

	Mi hermanito [My little brother]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-mi-hermano-chico
	Vaquita loca [Little crazy cow]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-la-vaquita-loca
	Cuncuna amarilla [Yellow caterpillar]	Mazapán	https://soundcloud.com/search?q=mazapan%20cuncuna%20amarilaa
	Mi Lapiz [My pencil]	PinPon	https://www.youtube.com/watch?v=rPxkpKcV7L8&frags=pl%2Cwn
	Alouette [Lark]	traditional French	https://soundcloud.com/une-vie-en-noir-et-blanc/alouette-gentille-alouette-1
	Ainsi font les petites marionettes [So dance the little puppets]	traditional French	https://soundcloud.com/jonathan-m-duchesne/ainsi-font-font-font
	La ronda de los amigos [Round of friends]	Mazapán	https://soundcloud.com/search?q=ronda%20de%20los%20amigos
	Una jirafa resfriada (A giraffe with a cold)	Mazapán	https://soundcloud.com/search?q=jirafa%20resfriada
	Francisca es una avispa [Francisca is a wasp]	Mazapán	https://soundcloud.com/sweetlauras/mazapan-francisca-la-avispa
	El payaso pelucón [The long-hair rabbit]	Mazapán	https://soundcloud.com/sweetlauras/mazapan-12-el-payaso-pelucon
	Mazamorra Del Poroto Coscorron [Dance of the naughty bean]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-mazamorra-del-poroto-coscorron
	Caracol Agustín [Agustín the Snail]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-caracol-agustin
	El globito [The little baloon]	Mazapán	https://soundcloud.com/search?q=mazapan%20el%20globito
	En una nube [In a cloud]	Mazapán	https://soundcloud.com/mazapan-music/en-una-nube
	Carnavalito del ciempiés [Carnival of the centipede]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-carnavalito-del-ciempies
	Remolino de papel [Paper swirl]	Mazapán	https://soundcloud.com/plaza-independencia-m-sica/mazapan-remolino-de-papel
Nursery rhymes or children's songs through which parents had previously interacted with their children (Inter Nurs)			
Participants 3-6	The wheels on the bus	traditional British	https://soundcloud.com/mothergooseclub/the-wheels-on-the-bus
Participants 1 and 2	If you're happy and you know it	traditional British	https://soundcloud.com/clovercroftkids/if-youre-happy-and-you-know-2

Pop songs through which parents had previously interacted with their children (Inter Pop)			
Participant 1	Beautiful things	Gungor	https://youtu.be/R7yyorwvuac
Participant 2	Man I feel like a woman	Shania Twain	https://youtu.be/E7auzP9RhCY
Participant 3	You can call me Al	Paul Simon	https://www.youtube.com/watch?v=YeYPLhCFrP0&frags=pl%2Cwn
Participant 4	I'm yours	Jason Mraz	https://www.youtube.com/watch?v=ThI9fm52lms&frags=pl%2Cwn
Participant 6	Africa	Toto	https://www.youtube.com/watch?v=wGjtv2jHKrg&frags=pl%2Cwn

Appendix J. Post-hoc interviews

1. Hugh’s mother

June 7, 2019. Skype video call.

[Debriefing about the study’s aim and results]

Researcher: so, for instance, if I may ask, at the time of the study, how familiarized was Hugh with screens? Was he particularly familiar with them? Did he have a tablet, or something like that?

Participant 1: um… no, he didn’t, really. His sisters do but he wasn't that… yeah we tended to try not to let him near screens, which I think is probably why he was quite fascinated with them because they were new and interesting.

R: I see… Um let me see if I have another question, um, how did you feel throughout this study? How was it like? Please be completely honest because, of course, it must have been daunting to be observed and so I tried to make it as casual as possible and so on, I tried to make like little Smiley faces and so on but how did it make sense to you, did they become natural at any point or… what's your appraisal after all of this time?

P1: Yeah I think after the first few time, yeah it was… it became not more natural, was easier, so yeah

R: so would you say that um, um it didn't feel completely artificial throughout?

P1: No, I don’t think it felt artificial throughout

R: um can I ask you a few more questions? So, you chose um this song you know the what is it what does the fox say what does the fox say what could you remember very briefly what was the story behind it?

[silence…]

R: If you remember….

P1: If I remember (chuckles)

R: yeah don't worry don't worry about it might come back. For Beautiful Things[1], what's the story behind that

P1: I don't quite remember, I really like that song but especially when I was pregnant with him that was a song I was listening to quite a lot, was quite relevant to me at the time when I was pregnant with him so it is always important song to me.

R: You chose ‘If you're happy and you know it’ for interacting with Hugh[2] did you hear that song as a child they did also have um a history dating back to your childhood was it something that your parents did with you, or something like that?

[1] Beautiful Things by Gungor was this participant’s choice of interactional pop (Inter Pop).

[2] [2] If you’re happy and you know it was this participant’s choice of interactional nursery rhyme (Inter Nurs).

P1: We did a lot when I was little, I remembered a lot of singing and because my mom was a music teacher, we would do a lot of music, and singing especially is something I remember doing a lot when I was a child and while being a pre-schooler, as well doing singing but yeah there was always singing at home as well.

[further conversation about the study and the possibility of being in touch for further studies]

2. Louisa's mother

June 11, 2019. Whatsapp voice call.

[Debriefing about the study's aim and results]

Researcher: okay so, um, a few a questions about some of the music you chose… so, you chose, for instance, Shania Twain's 'Man, I feel like a Woman', would you tell me briefly what's your story with this song, like why did you choose it?

Participant 2: basically, Shania Twain's my first big… basically I was a big fan of hers since I was twelve, and basically she's the one artist I collected all the CDs and… been a big fan since I was about 10.

R: OK, is t something your parents liked as well or an interest of yours, rather?

P2: no, and the first time I saw 'Man, I feel like a Woman', that I saw the video I was at somebody's house. I don't know why I got so hooked up [chuckles], sorry…

R: no, no, think she's fantastic! So, you also chose 'If you happy and you know it' for interacting with Louisa. Why did you choose that song?

P2: I like that one because it has actions as well, like 'clap your hands'.

R: similarly, did your parents sing that to you when you were a kid? That does … that song also take you back to your childhood?

P2: um… I can't remember them singing it, but it's one of the most popular ones, so obviously…

R: okay. So another question, um so I asked you whether you remember… I asked you to share… to interact with her on your own, like, without being observed, for some weeks, and then I played it during the sessions. Louisa's reactions to the song seemed way more enthusiastic than anything she had shown before to any of the other music, which is something that I expected. Did you have a good time interacting through it? Was there anything new for you in it, or for either of you, in in doing this? Or was it something that you were already doing when the study started?

P2: um, I hadn't listened to that song with her before, but we had done a lot of dancing in front of her before. I like dancing, so… she continues to dance [chuckles]

[the researcher offers to, should the participant be interested, send the footage over]

R: um so… so, sorry, I just want to get this right. So, you're saying that um you hadn't quite danced along this song before?

P2: yes, not to that song before, no.

R: okay, and that kind of interaction was new to her as well, in general, or just regarding the song?

P2: yes, just that song

R: okay, so you were already… you already had the experience of dancing and, you know, goofing around music.

P2: yeah

R: thank you for answering that. Looking at the videos I got the feeling… because my instruction was sort of 'don't prompt any particular reactions, but if Louisa has their own spontaneous reaction to the music, feel free to jump in'. During the first weeks, I got the feeling by looking at the at the videos that… you came across to me— I mean, this is of course a subjective judgement —, I felt that you were inhibited, that you might have felt inhibited in the beginning of the of the study, or is it just an impression?

P2: No, I'm quite nervous, so I was probably quite nervous the first few weeks…

R: Okay, yeah, I mean it was as if you… because she would look at you and smile and you smiled back, but somehow shyly, and I had the feeling by looking at the videos that she kind of… um, I don't know, you ended up smiling in spite of that, like she was so engaging and loving that I had the feeling that you couldn't help her at some point.. so that was quite nice, um, you'll see for yourself when I share the videos with you.

P2: [chuckles] yeah, yeah!

R: so, what is your honest recollection of participating in the study? What did you think of the setting? Do you have any considerations about it? Do you think it was natural or rather unnatural? Was there any… what is your like kind of recollection of that period… of participating in the study?

P2: It was more natural for her… not for me [chuckles]… because I was conscious that you were… watching… so…

R: anything else?

P2: I thank it was quite natural for Louisa, it was just me thinking 'someone's watching us'

R: so, for instance, if may I ask, at the time of the study, how familiarized was Louisa with screens at the time of the study? Was she particularly familiar with them? Did she have a tablet, or something like that?

P2: um… she got one when she was two, so she didn't have one. She'd probably be into our phones though, and laptop and stuff, she's good at looking for our phones [chuckles]

3. Daisy's mother

June 11, 2019. Skype video call.

[Debriefing about the study's aim and results]

R: In order to interact with her [Daisy] off-camera, you chose the song 'The wheels on the bus', right?

P3: yeah

R: um, does this song bring you back to your own childhood? Was it something that your parents sang to you, or something like that, or not really?

P3: no, I don't think so. It's just it's quite a common nursery rhyme. She still loves that nursery rhyme

R: [chuckles] oh, yes, I mean it was the most popular by far…
Okay so, thank you very much. How familiarised was Daisy with screens at the time of the study— this is roughly three months ago… Perhaps an easier question is: did she have a tablet of her own at the time?

P3: no… she does watch videos, and did watch videos on my phone… on Youtube…

R: …at that time?

P3: yeah, I think so, probably. I think she's been watching videos for a while, but she doesn't have her own

R: so here's an important question: during one of the videos— I think it was in one of the first videos —you invited Daisy to climb on your lap and listen to music from there. That is something she kind of intuitively do from then on. So, my question is, it looked to me like it was what happens in music groups, did she participated in any music groups at the time?

P3: Yeah we do a music group

R: so, just to be clear, that was something that was happening already by the time we started this study

P3: yes

R: and am I writing in thinking that posture you assumed was very similar to what would happen in a children's music group?

P3: yeah, yeah

R: so, I had children from different age groups, and from your age group Daisy was the one that paid the longest attentional bouts to music by far. I wanted to know this because in this sense she was a bit of an outlier, so I thought "if he hasn't been to a music groups I don't know how am I going to interpret these results', but now that you confirmed this I sort of relieved.

R: I have a few more questions. First, as I mentioned, some aspects of the situation I devised were useful (for instance Daisy couldn't see me or hear me, and so on), but still, what is your honest recollection from the sessions? How did they strike you? Did you feel it was any good? Did you feel it was awkward or kind of artificial to your daughter? What was your impression about to what extent it created a space for her to listen to music, or whatever?

P3: Well, it feel so long ago I can't remember exactly but, trying to recall, I feel like it was quite nice to spend time completely just together, like, obviously we did spend a lot of time together anyway but, you know, like purposely with each other doing the same thing. I think because we feel, like, if I remember one of the things is that we weren't supposed to be disrupted, so we always went to this spare room, so that was a bit like… that was unnatural

R: unnatural?

P3: yeah that was unnatural, because we wouldn't ever usually go and just sit in that room. Does that make sense?

R: of course, makes sense.

P3: so yeah, we would go in the bedroom, I would get the computer on and get it all ready and then sit down, and so on

R: that's all very helpful insight. Seems like none of this prevented her from listening to the music… or at least I didn't see her distressed

P3: oh no, not at all. I don't know like, say if we just sat in the living room and then we could just play the music it would have been maybe… I really, like, I can't remember now but she would like mess around with things in the bedroom didn't she?

R: sometimes, sometimes…

P3: Since we were in somewhere like other living room where we sit all the time she might have given it more attention because there was less things that interested her to pull around, because she can pull around all the time, but obviously that was my choice to do it in the bedroom you didn't say that I had to, so yeah…

[further conversation about the footage and the possibility of sharing it]

4. Patricia's mother

June 8, 2019. Phone call.

[Debriefing about the study's aim and results]

R: I'll start with some general questions. What is your honest recollection of the study? Did you enjoy it? Was it hard? What comes to your mind when looking back?

P5: that my daughter was stubborn when there was something different going on [chuckles]

R: right, that's very interesting. I was gonna' ask about that later on. Could you develop a bit more?

P5: it's just because… Normally, we'd listen to some music while in the car, or singing in the car, or singing around in the house when there's music on. But we don't really stay in one place listening to it, unless it was some kind of class, but even then there was some kind of interaction with other children

R: Absolutely. So, by the time of the study, had your daughter been in any music groups?

P5: yes, we have been to some music groups.

R: OK, so she knew what it was to… like be around you and listen to music and so on…

P5: yeah

R: and you're saying that there was something unusual about listening to music in that format or disposition

P5: yeah, in that kind of context, yeah

R: you say that it lacked the interaction…

P5: yeah, and in the music groups there would be colourful scarfs and stuff like that, and we would listen to nursery rhymes

R: I see. You mentioned the word 'stubborn'. How do you think the setting I provided impacted your daughter? Why did she behave in the way she did?

P5: just because I know that we listen to pop music all the time, and we still do, and… and she wouldn't do 'no, I don't want to listen to that, or that'… and I know she has a stubborn streak, and she was going through it at that time. It was her way or no way at all

R: I see. That's very interesting. It matches my own impressions. It seemed to me that your daughter kind of knew whether she was please or not by whatever she was listening to, so what you're telling me seems to match these impressions.

P5: yeah.

R: By the time of the study she was (…) around 24-25-months-old… was she in kind in this kind of 'stubborn mode' long before the study or was it something new…?

P5: no… she… started being stubborn around 20 months… and we're still going through it to some extent… but not to the degree that she was showing at the time. And, also, she's now able to verbalise a lot more clearly about what she's happy or not happy with, about what she wants…

R: perfect. This makes a lot of sense. You chose 'The wheels on the bus' as the song for interacting with her. Was there a particular reason for such a choice?

P5: um… just because… she likes all nursery rhymes but that was one she used to participate with, and do the movements…

R: … in one of these groups?

P5: yes, but also, at home, we have the musical books where you can press and it sings along. We have quite a few of those at home

R: and is 'The wheels on the bus' a song that you remember from your own childhood?

P5: [pause] I honestly can't remember. It was quite a long time ago [chuckles]

R: [chuckles] yeah, I perfectly understand. How familiarised was she at the time of the study. For instance, did she have a tablet of her own or something?

P5: so… she had… an Amazon Fire, and she used to listen to nursery rhymes on… in the morning as I was getting dressed. So she used to listen to nursery rhymes on YouTube. And then she uses the screen in order to choose the song.

R: I see. One last question: you mentioned that the setting was different from that of music groups you had both attended… do you think it had anything to do with the way she reacted to music, or would you say she reacted to music in the way she normally would?

P5: Not quite, sometimes. There were times when you got more around to her, also depending on mood she was in. I think I recollect times where she would interact completely throughout a song… so it depends on her mood, also the music choices.

R: and she was used to choosing her own music at the time (?)

P5: yes. Obviously not in the music group, but when it was just me and her

R: that makes a lot of sense. It seemed to me that she was able to pay attention quite autonomously, which is something younger children did not seem as able to do

P5: yeah

R: Was it something that was already there?

P5: Yeah.

R: would you like to add anything else?

P5: no, nothing.

[further information about potential future sharing of results]