Universität Bielefeld

Fakultät für Linguistik und Literaturwissenschaft

Dissertation

The collaborative construction of early multimodal input and its significance for language acquisition

12 MUTTATIONEN VON MALTE H.M. SPINDLER

von:

Iris Nomikou

Mai 2018

Universität Bielefeld

Fakultät für Linguistik und Literaturwissenschaft

Dissertation

zur Erlangung des Akademischen Grades

eines Doktors der Philosophie (Dr. Phil)

The collaborative construction of early multimodal input and its significance for language acquisition

von:

Iris Nomikou

Mai 2018

Erstgutachterin: Prof Dr. Katharina J. Rohlfing,

Universität Paderborn (ehem. Bielefeld)

Zweitgutachter: Dr. Akira Takada, Kyoto University

Acknowledgements

Arriving, finally, at the end of this journey I feel there have been so many people, moments, and discussions which inspired me, helped me find answers, made me reconsider, and supported me be it with chocolate, with wine, with coffee, a joke, or a simple hug. To them all I extend my sincerest thanks.

First of all to my supervisor PD Dr. Katharina J. Rohlfing, who guided and supported me throughout the entire process. I thank her not only for being a genuine "Doktormutter", in the most literal sense of the word, but for being all the different persons that I needed in every step of the way: a role model, a friend, a fellow traveller, a tap on the shoulder, a warm hug, good advice, a co-author, a neverending source of creativity, and the force pulling me back on the ground every now and then. Apart from being a mentor, she also helped me sharpen my eye and discover my own little voice. This will accompany me forever.

To my second supervisor Dr. Akira Takada for the sincere interest in my work right from the start and for his patience.

To my colleagues of the Research Group Emergentist Semantics for the endless discussions, their availability to any question, and their moral support. Special thanks go to Angela Grimminger for her personal involvement and support in any possible practical or personal issue, to Lars Schillingmann for the spontaneous technical assistance (nowadays via Skype) and for the

Contents

Foreword

If one's objective is to address the emergence of language, then considerations should commence by imagining how the linguistic environment may appear to someone without knowledge that language even exists, not to mention that language bears meaning. Born into our multisensory world, infants are faced with the task of somehow selecting from all the surrounding noise those aspects of the world, which are relevant for language, and they achieve this through experience with the environment. Yet, in this task they are not alone. Infants participate in language right from the beginning of their life. They are carried around, are spoken to and smiled at, they are tickled, and rocked to songs and music, they are shown things, and given things to touch, taste, and smell. The world, thus, impersonated as parents, siblings, friends and relatives, selects relevant information *for* infants' experience. This makes it so much easier for infants to find out what is important in our world. Infants are also immersed in meaning right from the beginning of their life, as they participate in everyday interactions. They smile and are smiled

at. They give and someone takes. This co-activity with others is meaningful, because that is what meaning is: sharing an experience with someone else, and pursuing a joint goal.

Two aspects of this interactive experience are central to the current study. The first is time.

> "Experience unfolds in time, in subjective as well as physical time" (Stern, 2000, p. 21).

What this quote indicates is that perception and experience must be viewed in terms of their duration or even recurrence, because experience is nothing more than the moment-to-moment live feed of sensations as they evolve and transform in time. Facial expressions, speech, gestures, actions, and even thoughts, all have durations, and if one is to perceive social signals such as language as they are embedded in the environment, then the ability to perceive and discriminate among temporal factors seems a judicious place to embark on theorizing about the onset of communication and language. Yet apart from the ability to discriminate among sensory cues, time and more specifically temporal duration also unites cues across modalities into a coherent whole. Different signals, such as visual, auditory, tactile, and kinesthetic, when occurring simultaneously, constitute a multisensory present, enabling the perception of unity and meaning in the environment.

The second aspect that is central to the current study is interactional time.

> "[C]hild development is shaped by a cumulative experience of a co-constructed interactive process between parent and child." (Hsu & Fogel, 2003b, p. 1061)

When contemplating how one learns via experience, another key concept is that of recurrence. In order to realize that something has been learned, one must be confronted with an event at least twice and recognize having already experienced it. As such, knowledge may originate from recurrence, and if time can be considered a dimension which unites and drives experience, then the recurrence of experience, might be a dimension which drives development. That is because the most salient feature of repetitiveness is its pervasiveness.

10

Repetitiveness causes us to tune in to ourselves and the people around us. Any movement persisted in and repetitive tends to become automatic. It is as though recurrent and repetitive patterns enable our minds to rest, whereas their violation needs our full attention. Accordingly, experiences which are repeated in time become event structures or "formats" (Bruner, 1983). Joint action becomes structured over time and thus more productive. A parent changing the infant's diaper for the first time might be overwhelmed by the complexity of the task. The infant having his or her diaper changed for the first time might be frightened. Only a few weeks later, as a result of the repeated interaction between parent and infant, the activity seems automatic. As the parent reaches for the fresh diaper, the infant might already be lifting up his or her legs to make room for the diaper. After six months, the parent may only have to say "diaper" and the infant will understand. Thus, repeated structures constitute the infant's first form of knowledge, allowing him or her to anticipate events and form expectancies and in time become conventionalized forms of communication, such as language.

This tuning in to interaction, and the active role of both caregiver and infant in fine-tuning it within the course of development towards language, is the focus of the present study.

1. Introduction

1.1. Grounding cognition and language development

How do children advance from infantile sensorimotor activities to form categories and develop the abstract and symbolic thought required for the acquisition and use of language? This question has been central to the study of cognitive development and cognition in general.

One position has been that infants enter this world with specific hypotheses concerning what language is, and the development of language is that of a *discovery* or *maturation*. These theories, which have been called 'inside-out', assume that the mechanism driving language acquisition is domain-specific, i.e., it occupies its own module in the brain and uses own mechanisms (e.g. Fodor, 1975). The environment in which infants live and the language they encounter in their interactions are considered as a 'trigger' that merely sets the system in motion to discover language. It then gets customized to the language spoken in the environment the infant lives in. This view has the main disadvantage that is does not allow for a transparency of the process of language development and is, thus, unfalsifiable. It serves as a mere description of the results of development and cannot answer why development proceeds in the way it does.

A second position has dealt with language development as a process of *construction* and not discovery. The so-called 'outside-in' theories assume that language exists outside the child, in the environment. Accordingly, language learning is solved by domain-general learning procedures, i.e., by the same procedures that allow the child to analyze the environment. This position focuses mainly on the processes by which children learn language. Most importantly, language

learning is a totally bottom-up process, no different from learning that takes place in other domains. Within this position there has been a view suggesting that language acquisition begins with low-level associative mechanisms. Young infants begin to acquire language by associating word-sounds with perceptual information (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006; Smith & Yu, 2008; Plunkett, 1997). This type of learning is said to precede symbolic mechanisms which allow infants to understand that words are not just sounds associated with percepts, but referential in nature. This replacement of one mechanism with another is said to occur either suddenly (Tomasello & Rakoczy, 2003) or gradually (Hollich et al., 2000). Woodward (2004) criticizes this view in that "first words are [considered] disembodied associates" (Woodward, 2004, p. 149). In her view, it is unlikely that infants ever begin by treating words as such.

In concert with this view the present thesis posits that the attribution of learning to different mechanisms creates a gap. How does the infant move from using one mechanism to using another? To answer this the present thesis will focus on the continuity of development and seek to propose a developmental pathway, drawing on two factors which could accompany language acquisition from the very start without having to be transformed. One is *action* and the other *interaction*. What is meant by action is that infants experience language immersed in everyday activities in which language *is* action as it is used in the service of achieving some practical goal. Thus, the infant can use action to make sense of language from the beginning (Woodward, 2004). What is meant by interaction is, that development does not take place in a social vacuum, but within a certain social context, i.e., within daily encounters with social partners who can affect the way infants act, interact, and learn language. Such a view sees language learning not as a mental process according to which infants need to discover that words are symbolic and bear meaning. Such a view sees language as a living system consisting of physical interactions between people and the environment over time. This view of cognition has been named the grounded view. The following sections shall elaborate on its assumptions and relevance for the view on language acquisition presented herein.

1.1.1. Perceptual Symbol Systems and Embodied Cognition

In recent years, research on mental processes, cognition in general, and language as a higher cognitive ability has been increasingly embracing what has been named the grounded view of cognition. The supporters of this view (e.g. Barsalou, 2008; Thelen, 2000; Lakoff & Johnson, 1999) propose that there is no higher order abstract cognition, and that mental behavior is based on the physical interactions between people and their environment.

They reject cognitivist views on cognition, such as the computer metaphor (Turing, 1948 in Copeland, 2004 Turing, 2004). Such views regarded the brain as a computing machine, consisting of computational modules which execute cognitive functions through algorithms. It was assumed that sensory systems introduce perceptual information from the environment into the system. This information is then combined, reorganized, and represented in the brain, transformed into abstract symbols. According to such representational views, perceptual and conceptual processes constitute two different systems which function on the basis of different principles. Perceptual states emerge in sensory-motor systems, and a subset of this perceptual information is subsequently transformed into a representation that is not perceptual. When viewing language as one instantiation of a system which operates this way then it becomes what Barsalou (1999) calls "amodal" and arbitrary since the way we represent language in our mind "bear[s] no correspondence to the perceptual states that produced" (Barsalou, 1999, p. 578) what we represent. Barsalou argues that in such systems, the perceptual information is necessarily transformed or "transduced into a completely new representational system" (Barsalou, 1999, p. 579). Instead, it seems more plausible to consider cognition to be grounded in the human body and its interaction with the environment.

> "Rather than viewing the body as a support system for a mind that needs to be fueled and transported, they view the mind as a support system that facilitates the functioning of the body." (Pecher & Zwaan, 2005, p. 1)

By acknowledging the body as the basis for mental behavior, these researchers assume that

mental processes are supported by the same processes used for physical interactions, that is, those used for perception and action.

When considering perception in Barsalou (1999) theory of Perceptual Symbol Systems, cognition is considered "inherently perceptual" (Barsalou, 1999, p. 577). Barsalou (1999) suggests that perception and cognition exist in a common representational system. His view assumes that perceptual states arise in sensory-motor systems. Through selective attention, part of what is perceived is extracted and stored in long-term memory. At a later point in time, this perceptual memory is retrievable and can function as a perceptual symbol. In contrast to "amodal" symbols theories, such as those discussed above, perceptual symbols are "modal and analogical" (Barsalou, 1999, p. 578), because they exist in the same system as the perceptual states from which they derive. Furthermore, Barsalou (1999) grounds these symbols in neural states of the brain that underlie perception. They are "record[s] of the neural activation that arises during perception" (Barsalou, 1999, p. 583). Most importantly, perceptual symbols are unconscious neural representations. Being an associative pattern of neuronal activation, they are less rigid than discrete symbols, but are of a dynamic nature. Grounded in sensory-motor states, they can operate on any aspect of bodily experience, even proprioception, and are therefore inherently multimodal. Such perceptual symbols are further enriched by accumulated experience, and related symbols are organized around a common frame, a simulator, that allows the system to retrieve an entity or event in its absence. Through the construction of simulators, Barsalou accommodates for symbolic functions otherwise represented in traditional theories by the notion of concepts or types, thus formulating a perceptual theory of knowledge capable of implementing symbolic systems, such as language.

When considering action, many researchers use the term 'embodied cognition'. According to the view opinion of Thelen (2000) "[c]ognition arises from bodily interactions with the world and is continually meshed with them." In an embodied framework, cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capabilities. These capabilities are inseparably linked and "together form the matrix within which reasoning,

memory, emotion, language, and all other aspects of mental life are embedded" (Thelen, 2000, p. 5).

Researchers embracing this view (e.g. Thelen, 2000; Chiel & Beer, 1997), suggest that cognition emerges from an embedded system. They are opposed to information-processing characterizations of cognition which hold that the world provides information to our senses (input); that this information is then processed by our central nervous system, which goes on to command the body to act in particular ways (output). The mind in input-output models mediates between the world and the body. In an embedded system on the other hand, world, body, and nervous system are embedded and coupled dynamic systems. The nervous system is a *dynamic system* embedded in and coupled with the body, and together they are embedded in, and coupled with, the environment (Chiel & Beer, 1997). Additionally, the term *system* suggests that the elements composing the system, apart from standing in some relation to one another, are affected by their participation in the system (Wilson, 2002). Finally, the term *dynamic* denotes that behavior emerges as a pattern resulting from cooperation of the system's components.

1.1.2. Embodied Cognition and Language

Taking up an embodied cognition view on language means that there is no mental "language of thought" residing in our brains (Fodor, 1975). No division needs to be made between an input/output system and a representational system in which concepts and meaning reside. Meaning is the outcome of interactions with the world (Glenberg & Gallese, 2012), and concepts are fed by the way we perceive the world and function with our bodies in it (Gallese & Lakoff, 2005). Gallese and Lakoff (2005) propose that this is possible because of what they call "neural exploitation" (Gallese & Lakoff, 2005, p. 456), which is the ability of sensory-motor processes in the brain to take on new roles while retaining their original functions. The authors suggest that our brain is multimodal and, accordingly, so is language because it uses the same structures as our sensory-motor system.

With respect to language development within a grounded cognition framework, the theory

of Perceptual Symbol systems (Barsalou, 1999) provides an interesting concept referred to as "variable embodiment". "Variable embodiment is the idea that a symbol's meaning reflects the physical system in which it is represented" (Barsalou, 1999, p. 598). This means that symbols are dynamic and adaptive to specific environments. According to this theory, infants may, during early development, use selective perception to integrate multimodal experiences in their memory, constructing simulators to represent entities and events and map these to their immediate world. This ability to construct simulators promotes the acquisition of knowledge at a preverbal level. Long before infants acquire and use language, they can simulate many aspects of their experience. This knowledge about the world helps them attach words to already existing simulators. Furthermore, because of the adaptive power of variable embodiment, as infants' sensorimotor system develops, so does their conceptual system (see also Smith & Gasser, 2005; Smith, 2005; Thelen, 2000). Smith and Gasser (2005) also underline the incremental character of development suggesting that "[i]nfants' early experiences are strongly ordered by the development of sensory systems and movement systems" (Smith & Gasser, 2005, p. 7) underlining the continuing character of development. Thus, the same basic form of conceptual representation exists from the outset of a human life and remains constant across development; a radically new form is unnecessary.

A similar argument, regarding the continuity of perceptual and conceptual abilities in infants, was raised by Mandler (1988, 1992) some years before Barsalou (1999) evolved his theory of Perceptual Symbol Systems. Mandler also believes that cognition emerges from early perceptual and motor schemas and argues against a transformation mechanism in infants, which is supposed to occur at some point of development and during which non-accessible sensorimotor schemas are qualitatively shifted to a higher form of representation, that is, a symbolic or conceptual system. Mandler (1988) believes that conceptual mechanisms are already in place from the beginning of human life. She proposes a type of "redescription" (Mandler, 2000, p. 19) of the information delivered by the perceptual input systems, which she calls perceptual analysis.

'Perceptual analysis' is "[a] symbolic process, probably conscious, by which one

perception is actively compared with another. This process can occur either simulta-neously (by comparing two objects with each other) or sequentially (by comparing a present object with a previously stored representation)." (Mandler, 1988, p. 126)

Perceptual analysis is attentive. Through perceptual analysis, information from perceptual input is extracted and recoded into a non-perceptual form that represents a meaning. This con-stitutes the earliest elementary proto-meanings from which the infants will develop concepts about the world they inhabit. Of course, perceptual analysis is likely to be quite primitive at first, especially in a newborn. Additionally, due to the immaturity of the central nervous system, the results of this process might not be accessible until after the first few months of life. Nev-ertheless, perceptual analysis might be considered an intermediate stop on the infant's journey from perception to language.

Further accounts of grounded cognition (e.g. Smith, 2005; Thelen, 2000; Fogel, 1993b) have emphasized bodily states, situated action, social interaction, and the role of the environment. This line of theory stresses the interaction between perception, action, the body, the environment, and other agents. The strength of this perspective is that it provides an explanation for infants' development of cognition and language. It seems more plausible to suggest that mental activity is grounded in continually perceiving the world and acting in it, because these are precisely the means infants have at their disposal. Drawing upon the above theory, it is suggested that infants are – right from birth – already in possession of all of the mechanisms which will drive their cognitive development and, consequently, their acquisition of language. It is both through the maturation of the nervous, sensory, and motor systems, as well as through infants' interactions with their environment, that cognitive development and language acquisition takes place.

In the next section, the mechanisms that could drive such an embodied language development will be discussed.

1.2. From perception to language

1.2.1. Ecological realism

Ecological psychology provides valuable insight into how infants acquire knowledge, which could be fruitful when asking how they learn language. This line of research emphasizes that the relation between a creature and its environment has consequences for behavior.

Gibson's ecological realism (J. J. Gibson, 1986) proposes the notion of affordances, referring to potential actions in the environment. Creatures whose bodies have particular effectivities or capabilities detect perceptual structures in the environment that provide opportunities for action. Depending on what a creature is capable of, it can discover what the environment offers as opportunities for action. According to ecological realism, creatures detect the perceptual structure that specifies the so-called invariants - i.e., stimulus properties that are somehow equivalent to each other - of ongoing objects or events, as well as the perceptual structure specifying transformation and change during cycles of perceiving and acting. Acting-perceiving cycles emerge in daily life. "As a person moves, affordances of the physical layout flow in and out of the perceptual field, some becoming prominent and others receding from view." (Zukow-Goldring, 1997, p. 202). Yet what is the value of this theory for language development?

The infant is born in a vivid, multi-sensory world full of sensations, in which he or she can perceive through a vast array of sensory systems: vision, audition, touch, smell, proprioception, and balance. J. J. Gibson (1966) suggests that invariants can be picked up across modalities from the perceptual array. This is called direct perception. For example, what looks like an apple, feels like an apple, and smells and tastes like one, will be perceived as belonging together constituting the object apple. This information is already structured in the environment, so the perceiver is not required to actively integrate it to achieve unity of perception.

For the development of language from an embodied view, this principle equips naive language learners with a valuable mechanism for perceiving and structuring the environment they inhabit. Infants are capable right from the start of their life of perceiving unitary objects, integrating

information from multiple sources, and thus constructing meaningful units which will in time become linguistic units.

1.2.2. Intersensory Redundancy Hypothesis

Inspired by J. J. Gibson (1966) view of invariant detection, Bahrick, Lickliter, and Flom (2004) and Bahrick and Lickliter (2000) propose an Intersensory Redundancy Hypothesis. This theory takes the principles put forward in ecological psychology a step further, by proposing how these can be and used as mechanisms of selective attention. In this view, information delivered to the senses from the environment is overlapping. One type of overlap involves amodal information, that is, information that is not specific to a single modality but is redundant across multiple senses. In contrast to modality-specific information, such as color or pitch, which can be detected by specific modalities, amodal information is not bound to one modality; for example the rate and rhythm of hands clapping can be perceived either visually or acoustically (unimodally) or via audio-visual overlap (multimodally). Bahrick and Lickliter (2000, p. 190) propose that the:

> "[d]etection of amodal information directs initial perceptual learning. It focuses the infant's attention on meaningful, unitary multimodal events (events in which stimulation from different modalities emanates from a single source) and at the same time serves as a buffer against the learning of incongruent or arbitrary intersensory relations."

Amodal stimuli include the dimensions of time, space, and intensity. For example, tempo, rhythm, or location can be derived from multiple modalities. The point raised by the Intersensory Redundancy Hypothesis is that such amodal information that is redundant over more than one modalities can be picked up by the infant's ability of invariance detection because of its increased salience and this has consequences for perception. Bahrick et al. (2004, p. 100) define Intersensory Redundancy (IR) as:

"[the] spatially coordinated and temporally synchronous presentation of the same information across two or more senses and is therefore possible only for amodal properties. (e.g., tempo, rhythm, duration, intensity)."

The Intersensory Redundancy Hypothesis consists of three components:

1. It acts on the level of infant attention, guiding selectivity and causing amodal properties of events to become "foreground" (Bahrick & Lickliter, 2000, p. 191).

2. This enhances perceptual processing, learning, and eventually memory for bimodally or multimodally specified properties before other properties, which are specified unimodally, such as color or shape.

3. The perceptual precedence of amodal information ensures unitary perception of single multimodal events and filters out further processing.

Thus, the infant's selective attention is guided to important events in the perceptual array. The filtering of other properties is initially considered a helpful mechanism of economy. Once infants have identified unified multimodal events, they can then increase their perceptual specificity by discovering other modality-specific properties (for the principle of increasing specificity, see E. J. Gibson, 1969). The Intersensory Redundancy Hypothesis also predicts that this mechanism should become less and less necessary throughout development, so that as infants grow, they will not need to rely on multimodal cues to detect objects and events in the environment. They will become able to detect perceptual cues as easily, if they are provided in only one modality.

Gogate and Bahrick (2001) expand the theory of intersensory facilitation in perception, by assuming that the same principles can apply to word comprehension and language acquisition, in general. This ability to abstract redundant information may contribute to later word comprehension. Possibly in some later stage, meaning is attached to such multimodal arbitrary pairings. Gogate and Bahrick (2001, p. 5) propose that:

"[t]he ability to detect redundant information across bimodal events not only pre-
cedes the detection of arbitrary relation but also guides the detection of arbitrary
relations within a given context."

1.2.3. Temporally Based Intersensory Relations

Lewkowicz and Kraebel (2004) employ a broader definition, according to which multisensory
redundancy is "the condition in which an object or event is specified by some combination
of amodal and modality-specific attributes" (Lewkowicz & Kraebel, 2004, p. 661). The au-
thors underline the role of synchrony in the integration of multisensory signals and speculate
that a different mechanism may underlie an integration based on synchrony than perception of
amodal invariance. Criticizing the Intersensory Redundancy Hypothesis (Bahrick et al., 2004),
Lewkowicz and Kraebel (2004, p. 661) argue that temporal synchrony is not a true amodal in-
variant.

"[...] temporal synchrony is an emergent property of two or more sensory inputs
being available at the same time, and thus cannot be perceived independently in
different sensory modalities."

Amodal invariants should be detectable even if they are not presented in a synchronous man-
ner, as in the case of cross-modal transfer, in which, for example, a rhythm is presented via the
auditory modality and subsequently the same rhythmical pattern is presented in the visual modal-
ity, and recognized. Thus, responsiveness to an amodal invariant inherent in a multisensory,
temporally synchronous event cannot not prove whether it originates from amodal invariance,
temporal synchrony, or both.

For Lewkowicz and Kraebel (2004) early in infant development, "the single most important
perceptual attribute that infants seem to rely on" (Lewkowicz & Kraebel, 2004, p. 661) is tem-
poral synchrony. Lewkowicz (2000) proposes a model with a focus on temporal relations such

as synchrony, rate, and rhythm, and suggests that temporal synchrony is the basis for the development of perception of more complex temporal intersensory relations. In what he calls the "epigenetic systems/limitations view" (Lewkowicz, 2000), temporal synchrony, rate, and rhythm are perceived successively and in that order with increasing infant age. This is owing to developmental limitations, and one type emerges from the other. Nevertheless, Lewkowicz and Kraebel (2004) emphasize that, even if temporal synchrony is driving perception, the facilitating and reinforcing role of amodal invariance should not be neglected.

1.2.4. Neurophysiological and neuroanatomical approaches

Behavioral evidence has also been documented by researchers interested in the cross-correlation of "time-locked" (Smith & Gasser, 2005, p. 16) multimodal experiences from a neurophysiological and neuroanatomical perspective. Some of these researchers suggest that temporal and invariant perceptual cues are biologically plausible (Thelen, 2000). Thelen and Smith (1994) review some findings in the neurophysiological literature and explain how the various senses, although to some degree localized in the brain, exist in "networks of interconnections" (Thelen & Smith, 1994, p. 189) linking anatomically distinct areas of the brain. They are convinced that a direct communication of the senses in the brain exists, such that they need not to integrated. Furthermore, they present evidence by Stein and Meredith (1993) showing that during visual processing, a large proportion of the neurons are responsive not only to visual but also to auditory and somatosensory input, and that the topographic maps of these modalities form "multisensory space[s]" (Thelen & Smith, 1994, p. 190) in the brain. They believe that during the detection of stimuli originating from multiple modalities, these interconnected areas of the brain proceed with "multimodal enhancement" (Thelen & Smith, 1994, p. 190).

1.2.5. "Acoustic Packaging" and the Coalition Model of language comprehension

This theory puts action in the heart of the language learning process, as it offers explanations for how infants might begin to link actions and their effects to verbal behavior and for how infants may construct their primary representations.

It proposes that the infant enters the linguistic world as a biased learner, sensitive to some aspects of the world over others. Similarly, as in the Intersensory Redundancy Hypothesis outlined above, this theory posits that the hypotheses the child may have about what might constitute language-relevant data are derived and constrained by the child's cognitive capabilities. The infant learns language using domain-general learning procedures, i.e., the same procedures that allow him or her to analyze the environment into ongoing events composed of actions and objects. As in the above described theories, learning is an entirely bottom-up process. Hirsh-Pasek and Golinkoff (1996) suggest that children use a coalition of multiple and overlapping cues in the linguistic and nonlinguistic environment.

The interesting aspect of their approach is that they see language acquisition not in isolation but within cognitive development as a whole. They propose that language acquisition and more specifically the segmentation of the speech stream, occur at the same time as the segmentation of the world, i.e., into objects, actions, events. Thus, these processes may assist each other. According to this theory, language learning is not the goal; the broader goal is cognitive development.

Accordingly, infants selectively attend to certain properties of language input and use multiple cues available in the input. They are assisted in this task by what the authors call a *primitive* use of language called 'acoustic packaging'. Hirsh-Pasek and Golinkoff (1996, p. 170) view acoustic packaging not just as an accompaniment or "attention getter" for the witnessed events, but also as one of many "carving knives". Thus, primitive forms of language comprehension in the form of acoustic packaging might assist children in "carving up the world" (Hirsh-Pasek & Golinkoff, 1996, p. 161).

In an early phase of development (approximately 0 to 9 months), the child uses the surface properties of language, such as prosody, overlapping with actions and events to represent objects, animate beings, and relations in the world. This is done by extracting unanalyzed "chunks" from the linguistic input that are associated with specific persons or events.

As mentioned above, language supports the development of action understanding by "assist[ing] the child in parsing the world's ebb and flow into units like objects, actions, or events" (Hirsh-Pasek & Golinkoff, 1996, p. 162). This may assist children in constructing primary representations: After the infants have extracted and internalized units and events in the world, they can begin to analyze within the acoustic units extracted and map words and phrases onto their corresponding representations of objects and events, leading them to the comprehension and production of language.

This model was further developed some years later as the Emergentist Coalition Model by Hollich et al. (2000, p. 18). It includes more input types, both perceptual and social, which could be used by infants learning language as, "children try to use all the pieces of information at their disposal, although not necessarily at the same time" (Hirsh-Pasek & Golinkoff, 1996, p. 51). The model is described below.

1.2.6. Emergentist coalition model

This line of theory is a hybrid approach, departing from the belief that children use multiple strategies to learn language. Specifically Hollich et al. (2000, p. 18) write:

1. Children are sensitive to multiple cues, attentional, social, and linguistic, in word learning.

2. Children differentially weigh certain cues over others in the course of word learning.

3. Principles of word learning are emergent as each principle changes from an immature to a mature state.

According to this theory, infants are biased to note certain kinds of inputs over others. For example, they are sensitive to attentional cues, such as perceptual salience, temporal contiguity,

and novelty. They are also sensitive to social cues, such as eye gaze, and can attend to social information from a very early age. Finally, infants are sensitive to linguistic cues, such as prosody and infant directed speech. Yet, the authors suggest that the ability to detect information is not the same as the ability to use this information to learn language. Their hypothesis is that "not all inputs are created equal" (Hollich et al., 2000, p. 23), meaning that at different stages of infants' development, the cues to which they have access are not interchangeable. Rather, they suggest that in the first months of life, infants make use of signals which cue their attention (e.g., perceptual salience, temporal contiguity, and novelty) to associate sensory information, for example simultaneously seeing something and hearing something, such as a spoken word and its referent (e.g., Gogate & Bahrick, 2001). Through experience, they become increasingly sensitive to other cues as they learn that, for example, pointing or eye gaze might be more a reliable cue for discovering what a word may refer to. Thus, as they get older, social cues are weighted more heavily than attentional cues. These are dynamic and emergent principles for word learning. Yet the combined action of all types of input is necessary for language learning. As the authors put it, what emerges is "more than the sum of each individual cue" (Hollich et al., 2000, p. 26).

This model has received some criticism. Although the authors claim that their goal is to demonstrate the process of change, Rohlfing (2013) points out that the model does not explain how an infant transitions from being perceptually driven to becoming one who attends to social information. Nevertheless, the value of this model lies in that it combines bottom-up theories of perceptual learning (like the Intersensory Redundancy Hypothesis) with socio-pragmatic theories, claiming that multiple factors from different domains must be taken into account and are equally important.

1.2.7. Summary

The theories presented so far speak for a perception-driven approach to language acquisition. At the heart of this approach are attentional processes allowing infants, with no top-down language-relevant knowledge, to select and attend to some aspects of the world which stand out, i.e, are

more salient than others. Guided by an embodied view on cognition, the mechanism behind this selective attention is considered biologically plausible, since the interaction of overlapping and/or correlated sensorimotor inputs resembles the functioning of the brain. Two aspects of these theories should be underlined here as they form the basic argument of this work.

The first has to do with the driving force of development, and it is the notion of the *education of attention*. In the ecological perspective on perceptual learning, the *education of attention* refers to the process of calibrating perception through interaction with the environment. Already in Gibson's theory (J. J. Gibson, 1966), the education of attention is central. It refers to the tuning of the senses to the environment a creature inhabits. Through the interaction with the environment, its structure is detected and affordances for action are discovered. This enables participation in new actions, which, in turn, enables the discovery of new affordances, etc. The same was suggested in the "epigenetic systems/limitations view" outlined above (Lewkowicz, 2000). To perceive rate, infants must first perceive temporal synchrony, thus their perception is educated by their experience. For language acquisition, this means that the perceptual systems detect language-world relationships and use those relationships to guide attention. This then leads to the detection of new relationships (e.g., Dent, 1990). In the Coalition Model of language comprehension (Hirsh-Pasek & Golinkoff, 1996), a similar strategy was suggested, namely that the perception of primitive multimodal language-action "chunks" educates infants' sensitivity to world-word mappings.

Yet the language-learning task involves much more than the detection of opportunities for action and the extraction of language-action chunks. It is inherently social. Infants engage in daily activities with caregivers, siblings, and other people who form part of their environment. Language, as a means of social communication, is thus part of the infants' environment. Observing this linguistic environment, one might discover the ways in which it is structured, enabling the infant to pick up linguistic invariants. In this case the invariants entail the relationship between the auditory structure in speech and a specific aspect of the perceptual structure (visual, tactile, olfactory, auditory, proprioceptive) available to him or her, allowing the infant to discover

how language is connected to the world and therefore learnable (e.g. Dent, 1990). For example, Zukow-Goldring (1997) proposes that development depends on the social-interactive environment of caregiving, as language emerges out of a socio-perceptual basis (Zukow-Goldring, 1990). She suggests that "[...] rather than perceptual structure flowing by infants like some uncut home movie, perhaps someone is editing the flow" (Zukow-Goldring, 1997, p. 205), thus promoting the noticing of affordances Zukow-Goldring (1990).

However, the education of attention within the social environment of the infant is not only about picking up perceptual signals, or cues, and associating them (e.g. Plunkett, 1997; Smith, Colunga, & Yoshida, 2010). Supporters of the socio-pragmatic view on language acquisition (e.g. Bruner, 1975, 1983; Akhtar & Tomasello, 2000; Mundy & Newell, 2007; Baldwin, 1995) suggest that the education of attention involves the perception of social-attentional cues, such as eye gaze and gestures, which regulate the coordination of attention within interpersonal interactions. According to this view, infants are assisted in their learning of language by their ability to perceive and follow these social signals which provide a hint towards what their interaction partner is referring to, thus "feed[ing them with] word-to-world mappings [...] in digestible portions" (Hollich et al., 2000, p. 12). This increases the likelihood of attending to the correct word referents in interactions. The position guiding the present work is, thus, that the education of attention is *social*.

The second aspect of the theories elaborated above concerns the mechanism of this social education of attention. Here, the suggestion put forward is that the education of attention consists of the interplay of language and action.

As already mentioned at the beginning, language and action are commonly viewed as two different skills: While the development of action understanding and production is considered to be a part of the motor system, language acquisition is linked to the development of perceptual, cognitive and social skills. However, when taking up a grounded, embodied approach on development, there are good reasons to claim that in children's development, language and action work tightly together. In the theories outlined above, it was suggested that very young infants

can make use of input originating from different modalities to help them perceive unity in the environment (Bahrick et al., 2004; Lewkowicz, 2000). This ability to pick up and integrate sensory signals is a low-level process, attributed to direct perception and infant's ability to detect invariants in the environment through intersensory facilitation (J. J. Gibson, 1966; Bahrick & Lickliter, 2000).

This is biologically plausible as according to theories of embodied cognition, unity of perception and dynamic categories for objects and actions emerge from the time-locked correlation of multisensory experiences that vary with the infant's activity and actions. Smith and Gasser (2005, p. 15) support this view by stating that "[a] multi-modal system that builds re-entrant maps from time-locked correlations only needs to be set in motion, to move about broadly, even randomly, to learn and through such exploration to discover both tasks and solutions."

The notion of re-entrant maps goes back to Edelman (1987) and his concepts of degeneracy and re-entry in neural structure. Smith and Gasser (2005, p. 4) define re-entry as "the explicit inter-relating of multiple simultaneous representations across modalities". Re-entrant maps are mappings between independent modalities correlated in real time. They allow the heterogeneous modalities to educate themselves and enable the system to "discover high-order regularities that transcend particular modalities" Smith and Gasser (2005, p. 5).

When thinking about language acquisition as a dynamic system, this notion is also helpful in thinking about how the child uses multiple sources of information, in this case sensory information, which is embedded in social interactions. Gogate, Walker-Andrews, and Bahrick (2001) suggest that, "[p]erceptual abilities in infancy dovetail with environmental stimulation to promote the development of lexical comprehension" (Gogate et al., 2001, p. 2). Turkewitz and McGuire (1978) suggested that the maternal role in child development should not be seen as restricted to emotional attachment or sensory stimulation. They propose, that a large part of her influence derives from the "integrated, multimodal stimulation" (Turkewitz & McGuire, 1978, p. 166). This ability, and the education of the ability, to abstract redundant information may contribute to later word comprehension. This is consistent with the idea of Sullivan and

Horowitz (1983), who proposed that the process of understanding words may result from realizing that spoken words refer, by convention, to perceivable objects or actions in the environment. Through their perceptual ability to detect intermodal invariants, infants pick-up information in patterns of stimulation "especially as their attention becomes educated by both experience and maternal scaffolding" (Gogate et al., 2001, p. 2).

Furthermore, it was suggested that language acquisition happens at the same time as action acquisition (Hirsh-Pasek & Golinkoff, 1996; Woodward, 2004; Rohlfing, Longo, & Bertenthal, 2012; Buresh, Woodward, & Brune, 2006). This speaks for a bilateral collaboration between language and action. Here, the social aspect of the input is deemed important. Yet, for example, in Hirsh-Pasek and Golinkoff (1996) no mention is made concerning what it is about the social input that makes it so important, apart from the fact that it is a source of information that the infants use. The argument which is raised by the present work is that the social input is especially powerful, as it consists of language, provided by more experienced members of the infant's social environment and within socially embedded action events. The action consists of the ways in which caregivers make events meaningful to their infants, by embodying their verbal messages with social cues, such as gaze and gestures, rendering them "tangible" (Zukow-Goldring, 1997, p. 206). Furthermore, the actions are the daily social interactions in which infants participate during which, according to Bruner (1975) for example, linguistic concepts and linguistic structure are first realized in ritualized repeated action.

Summarizing the arguments presented above, the suggestion put forward is that language emerges out of the socio-perceptual environment. Within social interactions, infants' attention is educated by the interplay of language and action and cued to pick up sensory and social signals, which guide them in discovering the relationship between words and referents and also in discovering how to communicate about something, by sharing a common focus of attention. The following two sections will elaborate on the role these perceptual and social signals play in language development.

1.3. Educating attention to social signals

The previous sections discussed that infants use a coalition of multiple cues, perceptual, social and linguistic in the process of language learning. Concerning social signals, a growing number of studies have discussed infants' perceiving of such signals, such as other people's gaze, and coordinating their gaze with an interacting person, in order to pursue a shared interactional goal (Tomasello, Carpenter, Call, Behne, & Moll, 2005). This ability, which has been called joint attention, enables the infant to not only look at either the interacting person or an object, but to look at an object together with the interacting person (Baldwin, 1995). This joint attention increases the infant's ability to learn from others. He or she can now perceive what the object of others' attention is and benefits from such episodes of joint attention (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998). A number of studies have shown how the perception and use of social signals enable children to participate in triadic interactions (Striano & Reid, 2006) and/or social frames (Trevarthen, 1979; Fogel, 1993b; Zukow-Goldring, 1996). Thus, long before infants utter their first words, they communicate by gaze (Jaffe, Stern, & Peery, 1973; Striano & Reid, 2006) and other contingent behavior (Kaye & Wells, 1980; Striano & Reid, 2006; Markova & Legerstee, 2006). This skill can be observed around 9-12 months of age, and has been recognized as "the 9-month revolution" (Tomasello & Rakoczy, 2003, p. 129), marking the unique development of communicative behavior in human infants.

The role of joint attention for the development of language is considered pivotal. First of all, because it allows interacting partners to share knowledge (Tomasello et al., 2005): When one person looks at something and the other person looks at the same object these two people share the experience of looking at the same object. Most importantly, though, it opens up a new dimension for communication as these two people are now communicating *about* the object they are both looking at. For the language-learning infant, this signifies the transition from learning that words are associated with referents to learning that the relationship between word and referent is symbolic, as the word is used to refer to an object, a property, or an action. Both aspects of learning are critical (Woodward, 2004).

The ability for joint attention was investigated in a study by Mundy et al. (2007). The authors identified four different types of behavior which are relevant for social coordination: responding to joint attention (RJA) is the ability to follow the direction of gaze and gestures of others, initiating joint attention (IJA) is the ability to use the direction of gaze and gestures to direct the attention of others, initiating behavior regulation/requests (IBR) is the ability to use gaze and gestures to elicit aid from a social partner to obtain an object or event, and responding to behavior requests (RBR) is the ability to correctly respond to a request. The difference between the joint attention measures (RJA, IJA) and the behavior regulation measures (IBR, RBR) is that in the former, the infants use gaze or gesture to refer to an object or follow someone's interest about an object. In the latter, the infants are requesting an object they cannot reach, or give assistance to a person. Interestingly, the authors found that only the joint attention measures (RJA, IJA) could predict the language skills of the infants at 24 months, thus providing evidence for the importance of this skill for language development. Yet, for the present work the question of interest is the course of development and the behavior leading to joint attention.

Recently, new methods have provided insights into the development of this skill. Gredebäck, Melinder, and Daum (2010), as well as Rohlfing et al. (2012), found that young infants demonstrate sensitivity to social signals, such as the congruence of pointing gestures, before they actually are able to point themselves. It thus seems that before joint attention can be observed in infants themselves, they have already become able to process functional aspects of it. One possible developmental pathway is that infants' perceptual and motor capabilities bootstrap joint attentional skills (Masataka, 2003; Rohlfing et al., 2012). Yet, there are different approaches discussing what drives the development on this pathway.

An alternative set of approaches explain joint attention as a result of developmental processes *within the infant*. According to a first account, the development is explained as the maturation of attention, cognitive, and affective systems as well as the interaction of these systems with children's own experiences (e.g., Mundy & Newell, 2007). A further line of research has proposed a theory, underlining the pedagogic significance of mutual attention (e.g., Csibra & Gergely,

2006, 2009). According to the theory of natural pedagogy, ostensive signals, such as eye contact or infant directed speech, allow infants to recognize that a particular behavior is specifically designed 'for' them and addressed 'to' them. This puts infants in a "pedagogical stance" (Csibra & Gergely, 2006), as it creates the expectation that what follows such a signal is something relevant for them; something they should attend to. According to this theory an education of attention is unnecessary, since sensitivity to social signals is considered as an evolutionary adaptation of the human species. This suggestion is supported by evidence showing that newborn infants show a preference, for example, for faces that appear to make eye contact with them (Farroni, Csibra, Simion, & Johnson, 2002; Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000). Thus, it is suggested that since young infants display receptivity to a signal which they have no former experience with, this sensitivity must be innate.

A further account suggests that joint attention is the result of a cognitive revolution that the child undergoes (Golinkoff & Hirsh-Pasek, 2006; Tomasello et al., 2005). It involves the awareness that other people have intentions. Also, in what has been called "secondary intersubjectivity", this revolution involves the infant'□□s awareness of the feelings shared in the relationship between the parent and infant (Trevarthen, 1993).

Finally, another possibility is that infants' perception is educated *within social interaction*. Within interaction, certain behaviors become established, as they are necessary in communication. This specializes or tunes infants to the resources required for communicating in a specific cultural setting and context of caregiving (Takada, 2011). Surely, this possibility does not necessarily exclude the others and may even augment them.

Socializing attention

This view echoes socio-cultural developmental theories which underline the role of ritualized, repeated interactions as the source of the necessary material for the child's construction of increasingly detailed scripts, or "formats" (Bruner, 1983) or "routines" (C. Snow, 1986). These interactions make regularities apparent for infants, as, according to Bruner (1983), they help infants

discover what to observe. The suggestion made is that language development and prelinguistic communicative development are embedded in social interactions (see also, Bullowa, 1979). This view emphasizes continuity in development.

For the development of mutual attention, Bakeman and Adamson (1984) outlined a gradual transition from interpersonal to object-focused attention under the influence of interaction. Accordingly, in the first months after birth, interpersonal attention involves dyadic (caregiver-infant) attention during face-to-face interactions that, not only give interacting partners contingent feedback, but also establish an interaction framework (Fogel, Garvey, Hsu, & West-Stroming, 2006). This framework is then transformed in the months that follow to encompass people and objects "outside the dyad" (Fogel et al., 2006, p. 41), through their gradual incorporation into the shared focus of attention.

For eye contact, as a first type of mutual attention, Grossmann, Parise, and Friederici (2010) carried out a neuroimaging study in which they measured 5-month-old infants' brain activation during eye contact and also when infants heard their name called (an arbitrary auditory signal which must be learned by the infant). They found that compared to adults, for which overlapping regions of the brain are activated for ostensive stimuli irrespective of modality (visual or auditory), infants recruited adjacent but non-overlapping regions (modality specific processing), which nevertheless might be functionally related. The authors suggest that this is a precursor for a later integration into a modality-independent representation of ostensive communicative signals. This finding provides evidence for the specialization process of the infant brain to new ostensive signals as these become established through post-natal experience with the social environment. It lends evidence in favor of the hypothesis that sensitivity to social signals develops in the first months of an infant's life and as a result of a calibration process (Csibra, 2010), which tunes infants' attention to perceive the stimuli which are provided by their specific environment.

Following this line of research, the hypothesis raised herein is that attention is educated in everyday social interactions. The suggestion is that the development of perceptual and social attentional skills is supported through reinforcement from the environment, because caregivers

"manage the communication process" (Colas, 1999, p. 114) early on by using attention mark-ers. This position holds that infants' (perceptual) learning is "nurtured"(Zukow-Goldring, 1997, p. 206) by caregivers who actively structure the input in systematic ways so as to guide infants in the detection of relevant features. Furthermore, maternal behavior is dynamically adapted to the infant's developmental stage (Gogate, Bahrick, & Watson, 2000). These approaches suggest that caregivers intuitively use infants' (pre)dispositions as a starting point on which they build up their own behavior and this, in turn, 'tunes' infants perception and action systems into their phys-ical and cultural environment. A special role is attributed to tutoring (Wood, Bruner, & Ross, 1976) in which infants' attention is directed (Zukow-Goldring, 2006) or constrained (Tulbert & Goodwin, 2011). The proposal is that the influence of the social environment operates on different timescales:

On a short-term timescale, it facilitates learning by providing structure in the input. This educates the infants' perception by actively influencing the mechanisms guiding it. On a long-term timescale, recurring instances or features of the provided structure lead to the emergence and stabilization of interaction frames or formats that become predictable, resulting in a long-term education of attention. Because this education naturally takes place within certain cultural contexts and is informed by the beliefs, values, norms, and specific goals of development, it is also a form of socialization of attention (Brown, 2011).

This "socialization of attention" (Call & Carpenter, 2002; Zukow-Goldring, 1990) is achieved by caregivers intuitively using infants' (pre)dispositions as a starting point to which they 'hook up' their own behavior. In this way, certain stimuli are reinforced and others are weakened, i.e., the infant's perception is narrowed down to the culturally relevant variation that the infant en-counters (Condon & Sander, 1974). This form of perceptual narrowing was shown for example in a cross-cultural comparative study, identifying culture-specific differences in the modalities of responsiveness emerging during the second and third month of life (Kärtner, Keller, & Yovsi, 2010). Comparing mother-infant interactions in Germany and Cameroon when the infants were between four- and twelve-weeks old, the authors found that, whereas maternal overall contin-

gent behavior showed no differences between cultures, over time there was a differentiation in the modalities with which this contingency was expressed. The German mothers used the visual modality more, whereas the mothers in Cameroon used touch and vestibular modalities. The interesting finding is that these behaviors did not differ from each other when the infants were four weeks old, but the distinct developmental trajectories became apparent with increasing infant age. The authors point out that the different developmental outcomes are the result of the culturally specific parenting beliefs that inform the patenting strategies of these mothers. This suggests that within the process of enculturation, certain resources—in this case mutual attention—are selectively reinforced and this enables the infant to adapt to the specific environment he or she inhabits.

The hypothesis raised herein is that this selective reinforcement consists of the interplay of language and action. The following section will provide some theoretical background to support this hypothesis.

1.4. Educating attention through intermodality

As previously formulated, the mechanism of this social education of attention was suggested as consisting of the interplay of language and action. Researchers in recent years have proposed that there is a close relation between action and language (e.g. Rizzolatti & Arbib, 1998; Zukow-Goldring, 2006). By relying on neuroanatomical and neuro-physiological evidence of so-called "mirror neurons" existing in primates and which become active both when an agent performs an action as well as when perceiving another agent performing an action, Rizzolatti and Arbib (1998) argue that human language may have evolved from this ability to link self-generated actions with actions of others, by simulating them in one's own motor system. By providing evidence from PET experiments, suggesting that a mirror neuron system also exists in humans, they hypothesize that this "observation/execution matching system provides a necessary bridge from 'doing' to 'communicating'" (Rizzolatti & Arbib, 1998, p. 188), explaining how a common

understanding, necessary for communication, may be established. The argument put forward is that acting and understanding one's own actions leads to understanding other people's actions, and that this in turn constitutes a basic level of understanding from which language may emerge. Further considerations concerned with the sensory-motor origins of language propose that language and action are of a common origin and are coupled systems (e.g. McNeill, 1992; Iverson & Thelen, 1999; Masataka, 2003). According to this theory, gesture and speech together operate on the basis of a common thought process. Given the proposed tight coupling of the vocal and motor system, the question arises, whether a coalition of language and action cues can support the education of attention. More specifically, there might be a symbiotic relationship according to which, action can support attention to language, and thus language development, and language can support the perception of action thus supporting the development of action understanding. The following section presents existing research findings supporting this symbiosis of language and action.

1.4.1. Action gives language a body

Naturalistic and experimental research has demonstrated that when language is coupled with synchronous actions, infants learn the presented verbal behavior better than when it is presented asynchronously to action. Zukow-Goldring (1997, 1996), for example, made observations in the home of mother-infant dyads, starting when the infants were 6 months old and throughout their one-word period, capturing the way in which caregivers "make previously unnoticed affordances perceivable to infants" (Zukow-Goldring, 1997, p. 211). In what she called "attention-gathering" (Zukow-Goldring, 1997, p. 220) interactions, she presented an account of how caregivers attract infants' attention to subsequently direct it to perceptual structure they have selected through "attention-directing" (Zukow-Goldring, 1997, p. 221) gestures, and suggests that language emerges through such interactions. She reports mothers' use of synchrony in the form of coordinated redundant information across more than one modality, such as tempo, rhythmicity, and accelerated intensity. Zukow-Goldring (1997) suggests that caregivers make events

meaningful to their infants, by embodying their verbal messages with actions, rendering them "tangible" (Zukow-Goldring, 1997, p. 206). As infants then see and hear events as they emerge, they may successfully establish the connection between speech and accompanying actions or objects, thus enabling the emergence of the lexicon. Furthermore, in order to detect a relation between words and world, the infants must perceive that "these two different kinds of things share an 'identity'" (Zukow-Goldring, 1997, p. 216). Caregivers thus engage in a social education of attention in every day activities and exploit multimodality of stimulation so as to select elements from the perceptual array, enhancing their saliency for the infant to detect. The role of the caregivers is deemed very important, as "[t]he mere co-presence of a perceptually sensitive infant and the availability of detectable information in the perceptual array does not guarantee that conventional relations will be detected" (Zukow-Goldring, 1990, p. 712). The way that caregivers support the learning infant is by limiting alternative interpretations (Bloom, 1995; Zukow-Goldring, 1990). "The problematic pairings are resolved through guidance by the caregiver" Zukow-Goldring (1990, p. 712), as caregivers build on the infants' behavior and "provide scaffolding to support it" (Smith & Gasser, 2005, p. 17).

In a more limited semistructured setting, Gogate et al. (2000) provide an account of what they call "multimodal motherese". They asked mothers to teach their infants four novel words by using four distinct objects. Two words were nouns and corresponded to two different objects. The other two were verbs and corresponded to two other objects, each of which had to be moved in a specific way to describe the action that the novel verb was to stand for. The authors found that mothers often presented the target word in temporal synchrony with object motion or touch. Embracing the Intersensory Redundancy Hypothesis (Bahrick et al., 2004), the authors suggest that this behavior highlights target word-referent relations for infants. This movement was less pronounced towards older children suggesting that especially young children need the "perceptual structure" as suggested by Zukow-Goldring (1996).

In further experiments, Gogate and colleagues found that when 6-month-olds (Gogate & Bahrick, 2001), and even 2-month-old infants (Gogate, Prince, & Matatyaho, 2009), were pro-

vided with a synchronous movement of the referent when hearing a syllable, they could detect changes in novel syllable-object pairings rather than when the referent moved asynchronously in time to the provided syllable. In addition, using eye-tracking technology with video, Rader and Zukow-Goldring (2010) presented videos, in which a person was showing a new object and labeling it in a (a) synchronous, (b) asynchronous or (c) static gesture condition. In the synchronous condition, at the time at which the object was labeled, the actor moved it forward (looming motion). In the asynchronous condition, the movement of the object began as soon as the actor started the script of labeling the object. Thus, in this condition, a movement was visible but it was not temporally synchronized with the label of the object. Finally, in the static gesture condition, the actor held the object and did not move it at all. The authors found that 9-to-15-month-old infants not only prefer looking at objects that were presented in a synchronous word-object condition, but also demonstrated a better comprehension of the word.

It seems that providing synchrony between action and language is a form of tutoring behavior occurring when addressing a child. Such tight intersensory coordination was also observed in studies analyzing caregiver input: Using a data corpus of semi-structured interactions of parents and their 8-to-30-month-old infants, Rolf, Hanheide, and Rohlfing (2009) implemented a computational model which uses audio-visual signal correlation and found that the multimodal coordination in adult-child interaction seems to be greater than in adult-adult directed interaction. In a similar vein, Schillingmann, Wrede, and Rohlfing (2009) implemented a model of acoustic packaging, using data from the same corpus with only 9-12 month-old-children. This model segments the movement stream with the help of acoustic signals, binding visual and acoustic information that overlap in time into a sequence. The authors also found that more acoustic packages were presented in child-directed than in adult-directed input. In addition, these audiovisual overlapping segments contained less content, because they contained fewer motion peaks.

Thus, synchronous action seems to be part of the intuitive parenting program (H. Papoušek & Papoušek, 1987; H. Papoušek, Papoušek, & Bornstein, 2002) and, as such, relevant for language

learning, but its effects can be even further extended: In another string of research evidence is provided suggesting that synchrony is also relevant for action learning, as will be elaborated in the following section.

1.4.2. Language disentangles complex actions, giving them a voice

The scaffolding role of language in the segmentation of action goes back to Hirsh-Pasek and Golinkoff (1996, p. 161), already presented in a previous section, who suggested "acoustic packaging" as a "more primitive use of language". What is meant by "primitive" is the fact that early in human development, language functions as a social signal rather than by its semantic content. This signal is a means to "divide a sequence of events into units" (Hirsh-Pasek & Golinkoff, 1996, p. 165). Hirsh-Pasek and Golinkoff (1996) proposed that by providing overlapping speech to some crucial parts of an ongoing event, some event boundaries will be marked, and these marked parts will be better perceived by infants.

From analysis of naturalistic data, Adamson (1995) suggests that mothers build "scaffolds" for their infants by acting like "narrators of the world surrounding them and their infants" (Adamson, 1995, p. 146). Brand and Tapscott (2007) attempted to experimentally test how narration overlapping with action might scaffold action processing. In a preferential looking study, 9.5-month-old infants were presented a sequence of actions (A-B-C) in the familiarization phase. In addition to the action, the language in form of a narration, like "Wow! Do you see what she's doing? She's blixing!", was provided and overlapped temporally with some actions (e.g., A-B or B-C). The authors found that in the test phase, infants considered those sequences of actions as belonging together that were "packaged" by the narration. More recently, Meyer, Hard, Brand, McGarvey, and Baldwin (2011) have shown that during action demonstrations mothers' speech is synchronous to their actions when interacting with 6-to-13-month-old infants. More specifically, when demonstrating how to perform some actions, mothers seem to align speech and related action (rather than to align speech in general with actions), providing a meaningful multimodal behavior (see also Gogate et al., 2000).

This is because language has the power to be perceived as a marker: Already newborns are sensitive to human speech (Vouloumanos & Werker, 2004)); infants can attend to their mothers'□□ voices even in the presence of noise (Bortfeld, Shaw, & Depowski, 2013). In addition, infants prefer human voices over other sounds (Muir & Field, 1979; Colombo & Bundy, 1983). This social signal not only plays a role in emotional attunement between the caregiver and the child (Markova & Legerstee, 2006), but also influences infants' perceptual categorization: Balaban and Waxman (1997) as well as Plunkett, Hu, and Cohen (2008) and Lupyan, Rakison, and Mc-Clelland (2007) have shown that when words are provided, infants (and adult learners) are more likely to detect crucial features in objects and categorize them. More recently, this facilitative role of language has been shown at a much earlier stage of development than assumed before: Ferry, Hespos, and Waxman (2010) examined the influence of words on object categorization in 3-to-4-month-olds and found that hearing labeling phrases, rather than hearing matched pure tones, influenced the infants' ability to form a category of the seen objects. Finally, even in neonates, a synchrony between maternal vocal activity and infant's body movements has been found (Condon & Sander, 1974) pointing to an early influence of sociobiological entrainment processes on the infant's sensorimotor system.

In sum, the reported studies provide convincing evidence suggesting that primitive meanings in event perception are formed when scaffolded with language input. On this early developmental level, infants are sensitive to language and can use it as a guide for their attention to redundantly specified parts of an ongoing event.

1.4.3. The symbiosis of language and action

Taken together, the evidence reported above strongly speaks for a symbiosis of language and action in which infants' understanding of actions is facilitated when packaged by language (operationalized as a vocal signal), and the rudimentary understanding of language is supported when accompanied by attention-directing action. At a perceptual level, when provided in synchrony in the input to infants, language and action reinforce each other, attracting and educating

the infants' attention (Bahrick et al., 2004). Here, synchrony refers to the fact that events in two modalities (such as hearable sound and visible movement) can be coordinated with each other temporally, and there appears to be a correlation between the characteristics of two events (Cangelosi et al., 2010). At a social attentional level, the symbiosis of language and action speci-fies socially relevant information *perceptually* for the infant (Zukow-Goldring, 1990). Through repeated interactions within everyday actions in which language is used, certain signals, such as eye contact or joint attention, become established as relevant and infants are thus socialized into perceiving them and using them. These interactions, apart from constituting the context in which attention is educated, are also driving forces of this education. According to interactionist approaches to language acquisition, learning within interactions is a reciprocal process of social coordination. Rączaszek-Leonardi, Nomikou, and Rohlfing (2013, p. 212) underline:

> "[the] crucial difference between acting that involves objects and acting that in-volves other members of the species. [...] actions specify parts of the environment: "the act of grasping is only comprehendible in relation to a thing that can be grasped" (Heft, 1989, p. 6). Similarly, one may state that the act of smiling is comprehendible in relation to a thing that can be smiled at".

A similar point is raised by Rohlfing and Deak (2013, p. 189) who suggest that when infants learn "they do not simply pick up information passively. They respon[d] to, and learn from, the interaction as they jointly determine its content and quality through real-time contingent and reciprocal coaction". The interactive nature of the education of attention, and what it affords for leaning will be addressed in the following section.

1.5. Language and action in interaction

As already addressed above, infants-to a great extent-do not learn in isolation. They learn within everyday social interactions. According to interactionist approaches, interaction is not seen as

the mere context in which development takes place. Interaction is seen as part of cognitive processes themselves (De Jaegher, Di Paolo, & Gallagher, 2010). Accordingly, to investigate learning within interactions means to abandon the idea of how individual cognitive mechanisms work. The mother or caregiver is not seen as a competent sender of information and the infant as a receiver. Learning rather takes place as the coordination between interacting parties "produces a net gain of information in the system" (Fogel & Garvey, 2007, p. 252). This is because in interaction learning is co-constructed, and "this allows individuals to bypass their own cognitive limitations" (Rochat, 2001, p. 139). For example, for the segmentation of perceptual events mentioned before in the section on acoustic packaging, Rochat (2001) actually suggests that interaction as such, in early protoconversations between mother and infant, provides scaffolds for the segmentation of perceptual events. Interaction is the source of shared experience out of which language emerges (Bruner, 1983; Tomasello et al., 2005).

This is achieved because in interaction continuous mutual adjustments are necessary. These coordination dynamics operate on two time-scales: On a first time scale they entail the moment-to-moment, real-time adjustments necessary to maintain an interaction going, a coupling of agents and world (Clark, 1999). These are not planned ahead of time, but are constructed online. On a second time-scale these dynamics act on multiple repeated interactions within a cumulative history of interactive experiences (Hsu & Fogel, 2003b). This is possible as "[e]ffects that explain minute quantities of variance in particular situations can account for final outcomes when those situations recur" (Bornstein et al., 1992, p. 818). On this time scale learning and development are seen as the result of the same process as forming human relationships (Rochat, 2001; Fogel et al., 2006). Considering the transition process from one timescale to another, Bruner (1983) suggests that it begins with social responses to the infant out of which a reciprocity develops, and a pattern of interaction that the infant anticipates. Bruner (1983, p. 28) formulates this idea in the following way:

> "In any case, a pattern of inborn initial social responses in the infant, elicited by
> a variety of effective signs from the mother—her heartbeat, the visual configura-

tion of her face and particularly her eyes, her characteristic smell, the sound and rhythms of her voice—is soon converted into a very complex joint anticipatory system that converts initial biological attachment between mother and child into something more subtle and more sensitive to individual idiosyncrasies and to forms of cultural practice."

Out of repetitive routinized interactive formats, a constrained, systematic, predictive setting is constructed within which the infant can learn to be communicatively effective without, and eventually with, language. But what is it about interaction that makes it a driving mechanism for development?

Some researchers have suggested that it is the contingent, exchange of actions, and the feedback provided by interacting partners. Keller (2000) suggests that for the infant to attend to a didactic situation, the activity of mother and infant must be attuned. This contingent reciprocity allows the infant to discover regularities between his or her behavior and the other person. The infant realizes that he or she can affect the other person as much as he or she can be affected by the other person. Furthermore, the feedback provided by the infant allows caregivers to flexibly shape and fine-tune the input.

In this approach, sometimes the *sameness* between the behavior of mother and infant has been underlined. It allows mother and infant to "share" subjective states and emotional expressions (Trevarthen, 1979). Mother and infant are affectively attuned (Stern, 1985), and can share feelings, actions and intentions: by simulating and recreating in one's own body the movement or sound of the other person, one can better perceive the movement itself and link it to possible intentions. Especially, in early interaction this skill forms an emotional link between an infant and a caregiver (for a review, see Trevarthen & Aitken, 2001). Rączaszek-Leonardi et al. (2013) speak of this process as a coupling mechanism, enabling mother and infant to enter an interaction and maintain or stabilize it, serving as a flexible "glue" between them (Rączaszek-Leonardi et al., 2013, p. 211). Accordingly, caregivers coordinate their behavior with the infant, with smiles, frowns, vocalizations, by mirroring it, reciprocating it (Trevarthen, 1979). More specifically, for

vocal development M. Papoušek and Papoušek (1989) have stressed the value of imitation. By imitating infant vocalizations, mothers "provide models which may most easily elicit matching responses" (M. Papoušek & Papoušek, 1989, p. 149). For language acquisition this acts as a reinforcement of particular sounds, carving infants' language production.

In other studies the *dialogical* aspect of interaction has been more highlighted. This approach underlines the value of the allocation of, sometimes, different roles in interaction, for the achievement of one overarching goal within a given situation (Fusaroli, Rączaszek-Leonardi, & Tylén, 2014). In such approaches, the complementary link between infants' behavior and mothers' responsiveness is highlighted (Hsu & Fogel, 2003a). Here, the behavior of one participant is seen as constraining that of the other (Rączaszek-Leonardi et al., 2013). More specifically, focus has been laid on the possible influence of caregiver responses to infants' initiatives within interaction sequences. For vocal development again, in a naturalistic settings in which mothers and infants were filmed during play, Hsu, Fogel, and Messinger (2001) found that infants produced more speech-like vocalizations when mothers were smiling and making eye contact with them. Also, Gros-Louis, West, Goldstein, and King (2006) showed that mothers provide not only contingent responses to vocalizations, but also respond differently to infant preverbal vocalizations with varying phonological properties. Apart from imitating infant vocalizations they provide more conversational-like responses, such as acknowledgements (e.g., *'oh really'?*) and questions. In an experimental setting, Goldstein, King, and West (2003) manipulated mothers' responses to infant vocalizations by instructing them when to respond. There were two groups, the contingent and the control group. Mothers in the contingent group were told to respond contingently to their infant's vocalizations with social signals such as smiling or touching. Mothers in the control group were told when to respond in a way that was not contingent. The authors found that infants who received contingent feedback vocalized in a developmentally more advanced way. This effect lasted also after the experiment when the maternal behavior was no longer being manipulated.

These results indicate that infants are sensitive to various forms of maternal feedback. Also

it highlights the fact that, within interaction, caregiver feedback is selective, as it is influenced also by infant feedback. In a study in which caregivers demonstrated actions to their 8-12-month-old infants, Pitsch, Vollmer, Rohlfing, Fritsch, and Wrede (2014) showed that infants' feedback in the form of, e.g., their eye gaze (signaling their attention or anticipating subsequent actions) modified the movements that caregivers made when demonstrating the actions, forming an interactional loop. This caregiver sensitivity has been greatly documented in the literature under the term *responsiveness*. Its role for development is elaborated below.

1.5.1. Maternal responsiveness and development

Responsiveness has been defined as "mothers prompt, contingent, and appropriate behaviors" (Bornstein & Tamis-LeMonda, 1989, p. 50). Researchers studying this aspect of caregiver behavior assume that certain infant experiences can have great influence over their cognitive and social development. Thus, responsiveness has been suggested to influence child development directly and indirectly (Bornstein & Tamis-Lemonda, 1997). On the one hand, by being responded to, infants learn that their behavior has an effect and this can be a motivator to learn (Watson, 1985). On the other hand, responsiveness leads to a secure emotional relationship with caregivers, and this can foster socio-emotional and, in turn, cognitive development (for a review, see Wolff & Ijzendoorn, 1997).

For language development maternal responsiveness is also deemed important. In a longitudinal seminal study with infants from 3-to-54 weeks of age, Bell and Ainsworth (1972) concluded that infants of more responsive mothers not only cried less, but were more likely to develop more varied modes of communication. It seems that the contingent response to an infant signal helps the infant assign a functional role to his or her behavior, which might support the development of conventionalized communication, leading eventually to language. With their responsive behavior mothers are seen as "gatekeepers" (Gros-Louis et al., 2006, p. 514) of what information is available to be learned, as they selectively reinforce certain aspects of the environment and certain behaviors of the infant (M. Papoušek & Papoušek, 1989; Zukow-Goldring,

1996; Goldstein et al., 2003). Also, by responding to infants' communicative signals mothers are establishing the same topic of interest as their infants. This establishes a common pragmatic frame for language learning (Bruner, 1983; Tomasello et al., 2005).

Research on maternal responsiveness as a predictor for language acquisition has mainly looked at maternal verbal responsiveness. In a longitudinal study, Tamis-LeMonda, Bornstein, and Baumwell (2001) tried to illuminate various dimensions of maternal responsiveness, by analyzing what mothers respond to as well as how they respond. They coded the different infant activities and the various vocal practices with which mothers responded to them, starting when the infants were 9 months old. They found that certain forms of responsiveness are more or less appropriate or effective at different developmental stages, suggesting that children play an active role in determining the input that they receive. So at 9 months, infants development was predicted by mothers' responding to children's object exploration by affirming their actions, labeling objects and events, describing and so on. The predictive validity of this type of responsiveness fades by 13 months. At that age, feedback to infants' vocalizations becomes a more powerful predictor, such as imitating and expanding vocalizations. This is consistent with the idea that at earlier periods of development mothers respond to behaviors that they consider communicative without them having to be intentionally communicative, such as facial expressions or crying, or gaze at the adult or at objects. As the child develops, the behaviors to which mothers respond are clearly communicative, such as verbal behavior, or gestures (Yoder, Warren, McCathren, & Leew, 1998).

In a further study, maternal responsiveness was related to infant joint attention skills (Rollins, 2003). The idea pursued by the study was to examine what the influence of the social environment is to infants who do not yet exhibit robust joint attentional skills, i.e., at 9 months of age. Contingent caregivers comments were predictive of language comprehension at 12 months, while joint attention episodes at 9 months were not. Moreover, maternal responsive behavior also related to language outcomes at 12, 18, and 30 months, even when joint attention was taken into account. This is important as it shows that the early linguistic environment relates to later

language skills.

1.5.2. Language and action in maternal responsiveness: Revisiting the symbiosis

As it has been suggested that children acquire language by relating language to familiar nonlinguistic forms (Bruner, 1975), some studies have looked at the nature of the nonverbal context in which children hear speech from their mothers. In a longitudinal study with infants at 6 and then 10 months of age, Harris, Jones, and Grant (1983) looked at topic changes initiated by the mothers and by the infants. They observed that, when initiating a topic themselves and not in response to the child, the mothers had a strong tendency to accompany their initiations with an action. For example, if the child had just picked up two bricks the mother would verbally respond to this by saying *"You're picking up two again are you?"* (Harris et al., 1983, p. 25). The authors suggest that when the infant provides the context for a topic, mothers only need to follow-in. Contrarily, if, for example, the mother was leading-in by opening a new topic, she would say *"Look I'm building up something for you."*, while making a tower with some bricks (Harris et al., 1983, p. 25). In this case the overlap of verbal and nonverbal context seems to be important. It seems, thus, that this overlap of verbal responsiveness and action seems to be part of the intuitive parental behavioral repertoire.

  In a follow up study with 16-month-old infants, Harris, Jones, Brookes, and Grant (1986) asked whether the rate at which young children acquire language is influenced by the whole social context of acquisition, not just parental language. Comparing two groups of children, one showing a normal rate of language development and one with a slower rate, they found a significant difference in the occasions in which mothers initiated verbal episodes with accompanying actions. Also, mothers of normal developing infants were better in *timing* their verbal behavior, so that it temporally overlaps with the infants' focus of attention. Thus, it seems that temporal synchrony seems to be important here. This was recently confirmed by a study in which it was shown that, in interactions with 9.5-month-old infants, only utterances which were both

semantically appropriate and *temporally linked* to infants' behavior related to infant expressive vocabulary at 18 months (McGillion et al., 2013). Harris et al. (1986) suggest that through these multimodal responsive episodes, typically developing infants have had more opportunities to discover consistent links between linguistic input and the non-verbal context, in which the input occurs.

In support of the view that redundant information across different modalities is particularly important early in infancy, Hsu and Fogel (2003a) examined whether maternal responsiveness to infant vocalizations consists of coordinated verbal and non-verbal actions. In a microgenetic study of early mother-infant interactions, they analyzed maternal non-verbal response types. They specifically coded for facial expressions, such as smile and a mock surprise face, touch, such as touching the infants body, vestibular (e.g., rocking or bouncing), and kinesthetic (e.g., bicycling of infant legs) movements. They found that mothers showed a preference either for unimodal verbal responses or multimodal responses, but not for unimodal non-verbal responses. This could be explained by the fact that the mothers were responding to infant vocalizations which may have been eliciting responsive behavior mainly in the same modality. The authors suggest that mother and infant form a complementarily linked whole, as the behavior of one *regulates* the behavior of the other.

The studies presented above open the discussion on the symbiotic relationship of verbal and non-verbal behavior discussed in the previous sections. By providing responsive behavior in multiple overlapping modalities, the mothers might be both constructing the online and contingent feedback, and, at the same time, making the input particularly salient, and thus noticeable, to their infants. As infants have been suggested to use a coalition of overlapping cues in the service of language learning (Hollich et al., 2000; Golinkoff & Hirsh-Pasek, 2006), this overlap of linguistic (maternal speech and prosody), perceptual (intermodal synchrony), and pragmatic cues (maternal responsive behavior embedded within the interactive context) might construct rich language learning experiences.

1.5.3. Responsiveness to infant gaze: Revisiting the education of attention

Apart from infant vocalizations, infant focus of attention has been a central aspect investigated in research on maternal responsiveness. Visual fixation communicates infant attention. Early in development it is an obligatory resource, whereas later in development it becomes optional as other resources become available to the infant, such as gesture or language (Bullowa, 1979). Filipi (2009) suggests that gaze is the infant's early "way of starting to do interaction" (Filipi, 2009, p. 3), as it allows the infant to take part in conversation and maintain his or her participation. In general, studies have shown a positive relationship between mothers' structuring of interactions to be responsive to the children's attentional focus and vocabulary development (Tomasello & Todd, 1983; Harris et al., 1986; Carpenter et al., 1998; Tomasello & Farrar, 1986). Hoff and Naigles (2002) suggested that a possible explanation for this might be the temporal contiguity between words and children's attention. Yet, most studies reporting a relationship of infants' focus of attention and language have looked at infants starting from 9 months of age. It seems reasonable, though, to consider maternal responsiveness to infant gaze, especially when looking at interactions with younger infants, as gaze is one of the earliest communicative skills infants acquire and is considered as the "first dyadic system in which both members have almost equal control over and facility with the same behavior" Stern (1974, p. 188).

Harris et al. (1983) focused on the nonverbal aspects of infants' behavior which is responded to by mothers, starting when the infants were 6 months old. They found that changes in infants' gaze were an aspect of infant behavior to which the mothers greatly responded. This strategy faded with time. When the infants were 10 months old, mothers reacted to gaze less and more to what their infants were doing. This suggests that responsiveness to gaze is particularly important in investigations of early infant interactions. For 3-month-old infants, Koester, Papoušek, and Papoušek (1989) examined the nature of temporal patterns in mother-infant interactions. The goal of the study was to address the question, whether mothers' non-vocal behavior has its own "pace-maker" (Koester et al., 1989, p. 144), or if it is dependent on the infant's attentional state. In a face-to-face setting, they observed mothers and their 3-month-old infants. They

coded tactile, kinesthetic, vestibular, and visual behavior of the mothers, as well as the infants' gaze behavior. They found that different modalities exhibited different rhythms, and that the modality chosen by the mother, and the tempo with which it was employed, were related to differences in infant signals of gaze responsiveness. The results were considered in terms of "intuitive parenting behaviors" (Koester et al., 1989, p. 152), providing infants with adaptive didactic support. Especially gaze at the mothers' face seems to be of special importance for interaction. For 5-month-old infants, Bornstein et al. (1992) found that infants' gaze at the mother was more responded to than infants' gaze at surroundings and objects. This finding, again, is consistent with the idea that at earlier periods of development mothers respond to behaviors that they consider communicative. (Yoder et al., 1998).

Accordingly, at a very early age, infant gaze is a first step in communicating as it expresses the infant's interest and availability for interaction (Filipi, 2009). In the first three months of an infant's life mutual attention is established through eye contact (Bakeman & Adamson, 1984). In these early interactions mothers are very responsive to their infants' gaze behavior and "fit" (Filipi, 2009, p. 3) their behavior to that of the infant (Fogel, 1977; Stern, 1974). According to Filipi (2009), this responsiveness to infant gaze reinforces this behavior of the infant, setting the foundation for protoconversations (Bateson, 1979; Bruner, 1983) teaching the infant its importance in interaction, thus "setting the stage for talk" (Filipi, 2009, p. 3).

Taking this under consideration, it then seems reasonable to consider maternal responsiveness to gaze, as a maternal strategy for recruiting and maintaining infant attention. Bornstein et al. (1992, p. 818) suggest that "[...] maternal responding serves notice to the infant that mother is attending [...] serves to *maintain infant attention* and *continue the interaction*",

Similarly, Hsu and Fogel (2003a) conclude that especially the coordination of verbal communicative behaviors with nonverbal communicative behaviors in maternal responsive behavior may be devised to capture infants' attention, while Harris et al. (1983) also suggest that the nonverbal context accompanying maternal responsive behavior is important, to focus the infant's attention appropriately. These studies raise the argument, developed in a previous section, on

the role of the caregiver in socializing the attention of infants. In this case, it is maternal responsiveness to infant gaze that becomes a reinforcer. In support of this argument, Keller and Gauda (1987) report of a study in which they associated early parent-infant eye contact with the quality of parenting. They found that parents who were tolerant and responsive to infants' gaze had "high gazers" (Keller & Gauda, 1987, p. 135), i.e., infants whose eye contact with one or both parents at three months exceeded the mean values of the sample. This suggests that responsive behavior of the mother motivates or educates infants to engage in mutual attention.

This section highlighted the role of the infant's interactive experiences, as a mechanism driving his or her development. It was suggested that interaction is not only the social context within which tutors or caregivers provide supportive input, but that the joint participation of mother and infant in interaction activates coordination dynamics, which are made evident in the contingent responding of one participant to the feedback of the other (De Jaegher et al., 2010; Fogel & Garvey, 2007). Within this framework, the behavior of one reciprocates (M. Papoušek & Papoušek, 1989) but also complements (Hsu et al., 2001; Gros-Louis et al., 2006) the behavior of the other. The supportive role of this interactivity for language development was illustrated in the various types of maternal sensitivity and their immediate influence on infants' prelinguistic vocalizations (Hsu & Fogel, 2003a; Goldstein et al., 2003), as well as their long-term influence as predictors of infants' prospective linguistic skills (Tamis-LeMonda et al., 2001; Tomasello & Todd, 1983; Harris et al., 1986; Carpenter et al., 1998; Tomasello & Farrar, 1986; Rollins, 2003).

Apart from verbal responsiveness (Tamis-LeMonda et al., 2001), focus was laid on the non-verbal aspects of maternal sensitivity and especially on the joint influence of coordinated vocal and non-vocal maternal responsive practices (Hsu & Fogel, 2003a; Harris et al., 1986). Furthermore, studies capturing maternal responsiveness to gaze were reviewed (Bornstein et al., 1992; Harris et al., 1983; Koester et al., 1989), and the importance of responsiveness to infant gaze especially in early mother-infant interactions.

The theoretical considerations elaborated in this chapter speak in support of an interactive

symbiosis of language and action. What remains to be clarified, is whether this symbiosis could prove to be a form of responsive behavior to early infant communicative signals. Furthermore, it remains to be clarified, if early forms of mutual attention emerging in the first half of the infant's first year, such as eye contact, could profit from this multimodal responsive behavior. Finally, it is still unknown, whether the multimodal reinforcement of early social signals could play a role in language development. These open questions motivated the study presented herein. The following section will define and motivate the hypotheses that guided its planning and execution.

2. Setting the goal and defining the hypotheses of the present study

Taking an embodied cognition view on language acquisition I suggest that in development infants use selective perception to integrate multimodal experiences in their memory to then map them onto their immediate world (Barsalou, 1999). Long before infants acquire and use language, conceptual mechanisms are already in place (Mandler, 1988), allowing them to form proto-meanings, from which they will later develop their concepts about the world as they learn language. In this developmental path from perception to language infants use various sources of information or cues which guide their learning: from salient perceptual information, the movement of their body and their interaction with the environment, to their engaging in social interactions with other humans; all this plays an important role (e.g. Smith, 2005; Thelen, 2000; Fogel, 1993b) as mental activity and language are grounded in continually perceiving the world and acting in it.

The focus of this study is the interaction of the infant with the environment, in the form of routine everyday caregiver-infant interactions, and the question how these can prove helpful for the acquisition of language. The suggestion put forward is that social interaction is a tuning device for the infants' senses. This tuning is a gradual collaborative experience of repeated interactions within which infants' perception is calibrated. Calibrated in the sense that through interaction the infant learns to notice those aspects of his or her experience in the world, which are relevant for the language which the infant will learn. In this view, theorizing on language

development should depart from specifying these tuning mechanisms by considering very young infants, long before they utter their first words, focusing on the continuity of development. For this reason, the present study captures infants' interactions starting at three months of age and accompanies their development until they are eight months old.

More specifically, the focus is on the infants' natural *linguistic environment*, which is the physical and social context in which language is embedded, embodied and used. Most importantly this environment is social; a world in which infants engage in daily activities with caregivers, siblings and other people, who use language, as a means of social communication. Thus, the interest lies in the detailed observation of mother-infant interactions that prepare infants for language.

Taking a social interactionist approach, the assumption is that in these interactions infants' perception is guided and thus educated towards certain aspects of the world which caregivers consider relevant for language learning. This education of attention is both towards picking up perceptual cues, such as sights and sounds belonging together or perceptually salient events, as also towards perceiving social cues, such as eye gaze and gestures. Here the various cues also support each other, such that the education of perceptual cues can be supported by social cues and the education of social cues can be facilitated by attentional cues. As Rohlfing (2013) points out, the attentional mechanisms relevant for language learning are interconnected like two gearwheels driving a system; one gear represents the perceptual and the other the social mechanisms. Accordingly, this study follows such an holistic converging-mechanisms and methods approach.

This approach allies with research proposing that development is guided by processes not residing *within the infant*, but by processes constructed *within interaction*. For the above defined education of attention, the argument raised is that the development of perceptual and social attentional skills is supported through reinforcement from the environment, by the active structuring of the input in systematic ways so as to guide infants in the detection of relevant features. If the attention of infants is educated, then one would expect to find some evidence of the influence of mothers on infants. Taking under consideration the above evidence the first hypothesis guiding

the present study is:

H1: Infant attention will be educated by the interaction. More specifically, there should be evidence of a relationship between infant gaze patterns and maternal gaze patterns. Also, there should be evidence of the development of this relationship as a result of the interaction history of the dyad.

Here, infant attention is operationalized as the infant's gaze at the mother's face. The first prediction resulting from this hypothesis is that there should be a relationship between the infant's gaze at the mother's face, and the mother's gaze at the infant's face. Furthermore, the interactional history of the dyad is operationalized through the comparison of mother-infant gaze patterns at two time points. The second prediction is that if mutual attention is affected by the interaction, there should be evidence of a convergence of mother-infant gaze behavior over time.

This education of attention is assumed to be the result of the active structuring of the input in systematic ways. As it has been suggested that caregivers use infants' predispositions upon which they 'hook' their behavior, the suggestion here is that caregivers will act in a way which fits the infants' own mechanisms of perceptual development. One mechanism suggested as driving infants' perceptual development was the symbiosis of language and action. It was suggested that, early in development, infants selectively attend to intermodally specified events presented redundantly, as this enhances their perceptual processing, learning, and eventually memory for multimodally specified properties (Bahrick & Lickliter, 2000; Bahrick et al., 2004). Also, it was suggested that this interplay of language and action, on the one hand, helps infants create a link between what words mean and what they refer to in the world, thus educating their attention for relevant aspects of speech. On the other hand, the symbiosis of language and action helps infants parse the world into meaningful events, thus educating their attention towards the understanding of actions. The most important perceptual attribute was discussed as being the temporal synchrony with which the temporally overlapping intersensory information is presented (Lewkowicz & Kraebel, 2004; Lewkowicz, 2000). Yet, up to the present, although many

experimental studies have shown infants' sensitivity to synchrony, there are not so many studies that have documented the symbiosis of language and action as it spontaneously emerges in everyday normal life situations. In addition, the existing naturalistic studies (Zukow-Goldring, 1996) or semi-structured studies and analyses (Gogate et al., 2000; Rolf et al., 2009; Schillingmann et al., 2009; Meyer et al., 2011) investigated infants in the second half of their first year, focusing mainly on hand gestures and hand movements and trajectories. This study aims at investigating the interactions which lead to providing synchrony appropriate to the infants' attentional skills at the documented age of the existing studies. More specifically, by documenting mother-infant interactions at as early as three months of infant age, this study looks at the input of mothers after the infants' "social smile revolution" (Rochat, 2001). Because at this age is considered as the interpersonal age (Bakeman & Adamson, 1984; Stern, 1974), in which infants are becoming more and more responsive to social signals (Lavelli & Fogel, 2005), one would expect to discover intermodal synchrony in various parts of the body, as for example the head or the face. Thus, the second hypothesis is:

H2: Even in early interactions mothers will use temporal synchrony between their bodily movements and vocal behavior to educate infant attention. There will be a systematic structuring of language and action in multiple modalities.

Firstly, temporal synchrony is operationalized as temporally coordinated maternal vocal behavior overlapping with body movements in various parts of the mother's body. According to the Intersensory Redundancy Hypothesis (Bahrick et al., 2004), the prediction is that for the younger infants a significant part of the interaction with their mother should be structured in a multimodal way. The second prediction is that, with infant development the vocal-motor synchrony in maternal stimulation will first increase. This is due to the fact that in the first six months infants become more responsive to social signals and, thus, to maternal input (Trevarthen, 1979; Lavelli & Fogel, 2005). This will provide positive feedback to the mother, who will in turn increase stimulation so as to prolong the duration of the sequences of positive affect.

Secondly, what is meant by systematic structuring is the way in which verbal behavior reinforces action, and the way in which bodily action reinforces the verbal signal. This is based on the expectation that the interplay of language and action, on the one hand, helps infants create a link between what words mean and what they refer to in the world, thus educating their attention for relevant aspects of speech (Zukow-Goldring, 1997). On the other hand, the symbiosis of language and action helps infants parse the world into meaningful events, thus educating their attention towards the understanding of actions (Hirsh-Pasek & Golinkoff, 1996). Here the prediction is that, even in the first half of the infants first year, the input provided by the mothers will be rich in such structure. Furthermore, the prediction is that with infant development there should be a difference in the structuring provided by the mother. As the infant moves from the interpersonal to the object-focused stage (Bakeman & Adamson, 1984; Stern, 1974), one would expect this structuring to change, so as to correspond to the infants' perceptual development. This change will then become evident in the modalities used to structure the input.

As for how the education of attention through synchrony might function, the argument is that in early mother-infant interaction mothers will selectively reinforce mutual attention. The suggestion is that the reinforcement takes the form of co-occurring behavioral modification. This educates infants to engage in mutual attention as it creates the expectation that when they do something different happens than when they do not. The co-occurring behavioral modification is of a kind which, as already suggested above, fits the infants' own mechanisms of perceptual development, i.e., it is the intermodal synchrony provided by the mother. This synchrony frames such episodes of attention, making them salient for the infant and making them rewarding for attention and memory. At a perceptual level, when provided in synchrony, language and action could then reinforce each other, attracting and educating the infants' attention. At a social attentional level through repeated interactions within everyday routines, certain signals, in this case mother-infant mutual attention, or infant gaze at the mother, could become established as relevant and anticipated by the infants, who would be thus socialized into perceiving them and using them. The third hypothesis guiding the present study, therefore, is:

H3: Mothers will use synchrony to reinforce infant gaze. The employment of synchrony will thus be coupled to the gaze behavior of the infant.

The hypothesis formulated above seeks to provide an account of how the education of attention develops in the first year of an infant's life, more specifically in the first half of the infant's first year, before the occurrence of joint attention. This raises questions about its course of development, and more specifically, about how both the gaze behavior and the multimodal input are shaped by the interaction and by the development of the infant. The underlying suggestion here is that there are two forces collaborating in development. First, there is maternal sensitivity and scaffolding and second, there are adaptive forces; the dynamics of the interaction.

Concerning the first force, the suggestion raised in the introduction is that caregivers have control over the communication process (Colas, 1999). They actively structure the input in systematic ways (Zukow-Goldring, 1997) by dynamically adapting their behavior to the infant's developmental stage (Gogate et al., 2000). Bruner (1983) stresses the active role of the caregiver for the development of the infant. His argument is that rather than caregivers being the "model" of a fully developed behavior for the infants to copy, their role is that of a sensitive partner to the child's progress, "fine-tuning" his or her behavior to the level of their child's ability. This is achieved because adults "arrange" the input (Bruner, 1983, p. 39). Thus, as the infant develops one would expect the mother to constantly modify her behavior, so as to guide the education of attention.

The second force shaping the development of attention is interaction itself. It is the context within which development takes place and the engine which drives it. Accordingly, developing within interactions is a reciprocal process of social coordination, in which mother and infant jointly determine its course through moment-to-moment contingent coaction. Through interactional loops they exchange actions and provide each other with feedback. As mothers respond contingently to infant signals they are providing input which is relevant for the infant, as is it coupled to his or her own initiatives. This then, in turn, motivates infants, educating their attention to what it is the mother is doing and establishing a common topic of interest. As infants respond

to the mother's initiatives, they control the input they receive and signal to the mother how to flexibly adjust the input to their growing abilities, by "weaving [her] own behavior around the child's natural activity patterns" (Smith, 2005, p. 17).

What follows this suggestion is that variability in the way in which these forces act on early interactions would be reflected in the later behavior of the infant. This assumption is consistent with socio-cultural theory and especially the work of Vygotsky (1978) who suggests that interaction characterizes development prospectively. It constitutes the "buds" or "flowers" of (language) development, rather than the "fruits" of development (Vygotsky, 1978, p. 88). The assumption here is that the guidance provided by the mothers in the way they structure the interactions to educate their infant's attention can push development in one direction or the other. According to Vygotsky (1978, p. 90):

> "Learning is not development; however properly organized learning results in mental development and sets in motion a variety of developmental processes that would be impossible."

According to this account, different qualities of mother-infant interactions will lead to different developmental outcomes (Keller & Gauda, 1987; Bornstein & Tamis-Lemonda, 1997). The prediction here is that the education of attention in individual infants, and mother-infant dyads, develops differently as a result of the differential weighting of the above-discussed factors, such as the employment of intermodal synchrony, maternal sensitivity, and the coordination dynamics.

Considering the bilateral collaboration of language and action (Bahrick et al., 2004; Cangelosi et al., 2010), it is plausible to argue that providing language as a social signal accompanying actions and events early in infancy might facilitate infants' action processing. A more elaborated understanding of action, in turn, might provide a robust conceptual basis for the meaning of words. Furthermore, language and action which, as hypothesized above, is coupled to the behavior of the infant constitutes input which is relevant for the infant as is it coupled to his or her

own initiatives. This way synchrony between language and action becomes more than salient multimodal stimulation, it becomes a form or maternal responsive behavior.

The benefits of maternal responsiveness for language were discussed in the introduction. It was shown that, for example, it affects children's prelinguistic vocalizations (Hsu & Fogel, 2003a; Goldstein et al., 2003), and predicts prospective infants' linguistic skills (Tamis-LeMonda et al., 2001; Tomasello & Todd, 1983; Harris et al., 1986; Carpenter et al., 1998; Tomasello & Farrar, 1986; Rollins, 2003). This strong association has been explained by way of certain characteristics of responsiveness which facilitate the acquisition of language. Tamis LeMonda, Kuchirko, and Song (2014) propose that responsiveness acts on two levels: At a social-interactive level, responsive input creates a pragmatic frame in which language is constructed out of shared attuned activities with caregivers; at a sensory level, responsive input is temporally contiguous and conceptually contingent to infant behavior, creating rich time-locked experiences. A similar argument was raised by McGillion et al. (2013) who reported that input that is both semantically appropriate and temporally linked to infant behavior at 9.5 months related to infant expressive vocabulary at 18 months. This view is quite consistent with the view presented herein on the temporal synchrony of language and action. However, existing studies have investigated the role of temporal contiguity in caregiver responses to infant behavior mostly in one modality, namely the verbal. Therefore, the suggestion put forward is to investigate whether temporally contiguous responsive input, which originates from multiple modalities and is provided in a temporally coordinated manner, could also be predictive of infants' language skills, as it would be matched to the infant's mechanisms of perceptual learning (Bahrick et al., 2004).

Furthermore, it is hypothesized that the infant has an active role in co-constructing the responsive input he or she receives. This is consistent with research suggesting that infants' communicative behavior elicits variable responses (Wu & Gros-Louis, 2014), and that infants' responses to caregivers elicit more responsiveness from them during routine interactions (Vallotton, 2009). This supports the suggestion that caregivers' responses to their children support infant communicative development by serving as contingent reinforcers for the infants' own commu-

nicative acts (Keller, Lohaus, Völker, Cappenberg, & Chasiotis, 1999). For this reason, it seems plausible that differences in infant behavior could also predict their language skills. Taking into account the above arguments, following hypothesis was put forward:

> **H4: The variability in the infant's attention to the mother, and the variability in the mother's responsiveness in early interactions, will be associated with the variability in the child's language acquisition and, especially, with the variability in the acquisition of verbs.**

Infants attention to the mother is operationalized as the extent to which infants attend to the multimodal input provided by the mother. Mothers' responsiveness is operationalized as the extent to which mothers provide their multimodal input according to the infants' attention. Furthermore, as synchrony has been suggested as involving the structuring of actions, it seems reasonable to assume that the effect of responsive synchrony should especially support the acquisition of words for actions, namely verbs. To test his hypothesis a parental report is used as a measure of the children's verb and overall language skills.

These considerations and hypotheses motivated and guided the conception, design, and implementation of the present study. This will be described in the following chapter.

3. Method

3.1. Design

In order to capture the nature of early mother-infant interactions, the study took place within a naturalistic setting. The aim was to collect ecologically valid data outside laboratory conditions. For this reason the study was carried out in the families' homes. The activity for the interaction was meant to be both age appropriate and familiar to mother and infant, drawn from their everyday activities. For this reason, mother and infant were observed during diaper changing. In addition, lying down on the changing table is a natural position, giving the infant the opportunity to control her or his own movement independently without needing to be supported by the mother. This is important, as the infant was free to direct his or her attention to whatever he or she wanted and was free to engage or disengage from the interaction with the mother.

For the purpose of the study, a foldable changing table was brought and set up. It was placed in the center of a spacious room, either the living room or the child's room. This was necessary to facilitate filming and because not all families had access to a changing table and filming space. In addition, using this changing table, all mothers acted not only in a comparable body position but also within a comparable space. Thus, by providing the changing table, the interaction environment in all households was standardized.

Data collection started when the infants were about three months old. By this age infants have passed the "two month transition" (Lavelli & Fogel, 2005), which has been suggested as a period of developmental reorganization in the patterns of infants' attention. Furthermore, since

it was of interest to capture the development of these interactions, a longitudinal design was chosen. Through repeated observations, one can track "developmental history as it is evolving" (Hsu & Fogel, 2003b, p. 1064). Recording was carried out every four weeks until the infants' eighth month of age prior to the emergence of joint attention.

The hypotheses elaborated in the previous section made video recording necessary as a method. Two HD video cameras (Canon HF10; Sony 3CMOS HDV 1080i) were used for video recording. They were mounted on camera stands. One camera was positioned behind the changing table (opposite to where the mother was standing), filming from the bottom and slightly offset from the center. This camera was directed toward the mother's face and upper body. The second camera was set up in front of the table behind the mother, thus filming on a higher level over her shoulder. This camera captured the mother's arms and the infant's body from the side, making her or his hands, body, face and gaze visible (see Figure 3.1). An external microphone was mounted on one of the cameras in order to guarantee high-quality audio recording. The second camera recorded sound via its built-in microphone.

Figure 3.1.: Camera views

The recordings took place at moments which the mother considered as appropriate. This was

normally after the infant had slept and had been fed. The appointments were normally agreed on some weeks before, but the mothers were encouraged to call the experimenter on the day of the filming in case they wanted to have the filming earlier or later. This was more the case for the first two sessions as the infants did not yet seem to have a fixed schedule, but with time the mothers also became more precise when agreeing on the time, since the infants seemed to settle into a rhythm the mothers could rely on.

3.1.1. Recruitment

Recruitment was carried out through a number of different steps. Initially, a press release was published in a local newspaper and broadcasted on the radio, and information leaflets were left in various facilities in which courses for mothers and infants take place. Also, a list of all locations offering such courses were compiled and midwives were personally contacted and informed about the study. Through these midwives, it was possible to access such courses and address the mothers personally, inform them about the study, get the contact information from interested mothers, and also leave flyers for mothers to contact me themselves. Also, informational material was left at another infant lab at Bielefeld University. Upon participation in the study of this other lab, interested mothers were informed of the present study and could then get in contact.

3.1.2. Subjects

Nineteen mothers and their infants initially agreed to participate in the study. All lived in the greater Bielefeld area (North Rhine-Westphalia, Germany). One mother terminated her participation after the second visit, as she felt that the study was too time-consuming. The data from this first visit were excluded from analysis. A second mother-infant pair missed two out of six visits. Since both visits were those chosen at a later step for analysis, the data were excluded from the sample. A third mother and her infant missed one visit.

Infants

All infants were born between December 2009 and April 2010. Out of 11 boys and 8 girls, 14 were firstborn, (7 boys and 7 girls). All had been born with no complications and were of average healthy birth weight. One infant was born with a slight hearing impairment, which was detected directly upon birth, and has been wearing a hearing aid since the age of 2 months allowing her to hear normally. One further infant was born with a developmental dysplasia of the hip (DDH) and was temporarily wearing a Pavlik harness. Nevertheless, during the recording sessions, that is, the diaper changing, the harness was always removed so that the infant could move freely just like all the other infants. The age of the infants at the time of the recordings can be taken from Table 3.1.

Table 3.1.: Infant age at each visit

Visit No.	Mean AGE (Months ; Days)	*SD*
1	3;12	5
2	4;7	6
3	5;5	6
4	6;4	7
5	7;2	6
6	8;0	6

Mothers

The average age of the mothers was 32 years (ranging from 24 to 40). All mothers had finished secondary school, 50% had some professional training, and 50% had an academic degree. All mothers were native speakers of German and spoke German to their infants. During the recording period, all but one of the mothers were on parental leave. Yet the mother who had returned to work took her infant with her to the office.

3.1.3. Procedure

As previously mentioned, the families were visited in their homes, except for one mother-infant dyad. This particular mother normally changed her infant in a playroom next to her office at work. Since it was a familiar environment for both mother and infant, recordings took place in this room following the usual procedure. (i.e., the changing table was provided as with all other families).

At the beginning of each session, the table was set up. It was always positioned in the same location. Then the cameras were set and adjusted. Their position was not changed for the remainder of the session. The mother was asked whether she and the infant were ready to begin. When they felt ready, the cameras were turned on and the experimenter invited them to approach the table. The mother was asked to change the infant's diaper *"as she normally did, when she was☐n't in a hurry"*. The experimenter then left the room to make sure that mother and infant were not distracted during their interaction. Mothers were encouraged to bring into the interaction any game, song, or other diaper-changing routine they normally used. The experimenter provided no toys. However, because many objects (such as diapers, wet tissues, lotion, clothing, toys, etc.) were a natural part of the interaction within this situation, mothers were allowed to bring along and use any object they wanted if these were a part of the routine.

Mothers were always assured that they could pause or stop at anytime they felt necessary. During the entire study, the recording was interrupted and terminated on only one occasion due to infant fussiness. A new appointment was made two days later and the recording was repeated.

As soon as the task was completed, the experimenter was called back into the room and had a short discussion about the procedure with the mother. Then the next appointment was scheduled. Throughout the entire study, the visits took place every 28 days on average ($SD= 1$). On the second, fourth and sixth visit, small presents were given to the mother for the infant to express gratitude for their participation.

3.1.4. Data recording

Recording was carried out at 25 frames per second. Out of a total 108 recordings, 3 recordings were missed because the family was unavailable and one recording could not be analyzed because of a technical problem with one of the cameras. These recordings amount to about 12.5 hours of video data, and about 43 minutes of data per mother-infant dyad. The durations of the recordings per visit are listed in Table 3.2.

Table 3.2.: Duration of recordings

Visit No.	Mean recording duration (sec)	*SD*
1	449	233
2	471	256
3	457	283
4	481	287
5	429	255
6	351	207

The recordings from the two cameras were synchronized and rendered in a split-screen view. The sound recording from the external microphone was used. Films were cut and edited with Final Cut Pro®film editing software. During editing, the start and end of the each interaction was marked. The moment at which the mother stopped talking with the experimenter and addressed the infant was marked as the beginning of an interaction. The mothers usually said something like *"So, let's start"* or *"Where are we?"*. This normally coincided with the beginning of undressing the infant. The point at which the mother lifted the infant from the changing table into her arms, or when she explicitly called the experimenter back into the room, usually by saying something like, *"We are finished"*, was marked as the end of the interaction.

3.2. Methodological approach

The methodological approach selected is a converging-measures approach, in which multiple methods of data analysis are combined, both qualitative and quantitative. This decision was derived by the research questions and hypotheses guiding this study. As Valsiner (2000) suggests, the methods employed for data analysis must be adequate with respect to the phenomena in question. Accordingly, there is one general methodological plan guiding any inquiry, and qualitative and quantitative data are two kinds of data representations. Which kind of data translation is used, depends on what questions one is asking and the methods used become different routes to the same end. Demuth and Mey (2015) see the systematic integration of different procedures as taking different perspectives in an attempt to provide a holistic description of a phenomenon of interest, while Fogel et al. (2006) posit that multiple methods may complement each other.

For this work, the combination of qualitative and quantitative analyses was necessary within the process of generating categories to describe the data, of generating hypotheses grounded in the data, and then proceeding to test these hypotheses. According to Demuth and Mey (2015), qualitative paradigms have a theory-generating logic, avoiding hypotheses made beforehand, and proceed in a data-driven manner, avoiding preformed closed categories. Any analysis following such a method preserves the connection between the described behavior and the context within which it occurs.

In the methodological approach of the present study, analyses starts with the collection and description of single cases of interest. Case analyses are of significant help in the search for patterns within phenomena.

> "To find out what people do in general, we must discover what each person does
> in particular, then determine what, if anything, these particulars have in common."
> (Thorngate, 1986, p. 75, quoted in Fogel et al. (2006))

In the above quote, what is meant with *"in particular"* is not only the idiosyncratic details of a single case, but also the situated meaning of behavior as this is observed within this case. By

analyzing single cases as they occur within interaction, within a certain situation, and between two specific people with specific interactional histories, one is gifted with rich contextual information which point to the meaning of the participants' actions. Rogoff et al. (1993) believe that because people provide explicit evidence regarding the meaning of their actions in interaction, so as to facilitate understanding, this evidence is made available to the researcher analyzing a specific case in depth. In this way, the collection of single cases allows the researcher to identify interesting instances and patterns within the data, along with the microanalytical analysis of change over time, and, hence, the study of developmental processes (Demuth & Mey, 2015). Also, Fogel et al. (2006) propose that cases can serve as hypotheses to be tested on larger data samples.

In the present approach this idea of hypothesis generation and testing was pursued. Yet, for the single cases to be transformed into larger data samples, the particular details of each case must be abstracted. Rogoff et al. (1993) characterize this process as one of "developing a common language for discussing" across cases (Rogoff et al., 1993, p. 29). This involved the development of categories to account for all cases and included the development of a coding scheme as well as the decision of how to segment the observations, and finally the microanalytical processing of the entire data corpus, using frame-to-frame analysis to segment and code the data according to the developed categories. This step requires the events to be broken down into discrete codes, thus decomposing the holistic qualitative description of behavior, making it seem isolated from its context and thus meaningless. Fogel et al. (2006) underline that this step is not theory but merely technique. As a result of this analytical step, the observations became measurable (Bakeman & Quera, 2011) and this allowed for the expansion of observations from single cases to all cases in the data.

The next step involves condensing the data and making patterns visible. For this, statistical analyses were used. These enabled the tracking of observations across multiple variables and the visualization of possible patterns of interrelations among them, thus allowing a generalization of observations and the testing of hypotheses which were derived in previous stages of the

observational analysis.
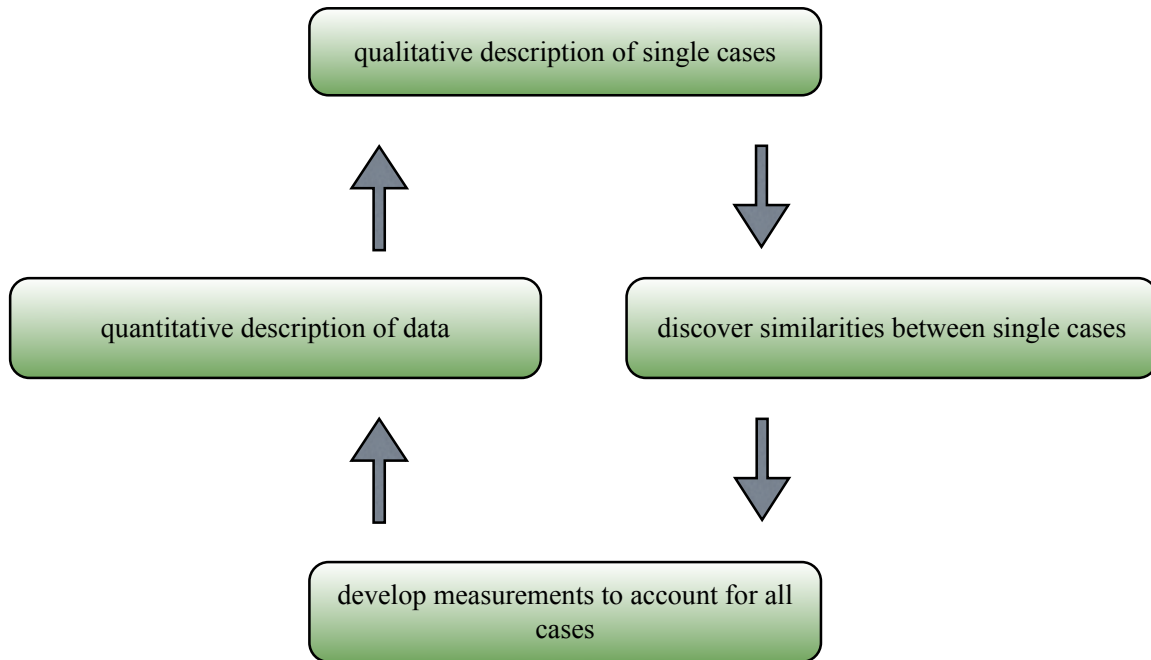
Figure 3.2.: Methodological loop guiding the study

Altogether this approach made it possible to zoom into the data and observe behavior as it occurred within a concrete situation or "contextual configuration" (Goodwin, 2000, p. 1490), and collect single instances of a phenomenon. The next step was to zoom out and look for similarities between these instances, which could in turn be generalized to the entire sample. Most importantly though, this approach forms a methodological loop (see Figure 3.2), allowing the analyst to return to the raw data to check whether generalized statements were indeed grounded in the individual cases. This way the behavior was never decontextualized.

This part outlined the general methodological considerations guiding the analyses of the study. The specific implementations of this approach and the analytical tools used will be elaborated in each of the subsequent chapters investigating the formulated hypotheses.

4. Mother-infant gaze patterns: a longitudinal comparison

4.1. Objective

In the introduction, the argument was raised that infants' social environment affects their attention very early in development in order to establish a condition of mutual gaze. This is because the interaction between infant and caregiver is empowered by mutual attention, and this manifests itself in the establishment of eye contact. Eye contact is an important "interactional topic" in the first months of life (Keller & Gauda, 1987, p. 130), which opens the channel for communication (Farran, Hirschbiel, & Jay, 1980). Keller and Gauda (1987) suggest that as such it reflects the quality of the interaction. Most importantly, it could be predictive of social and cognitive outcomes later in development (Keller & Gauda, 1987) and a potential "forerunner" of other forms of communication (Messer & Vietze, 1984, p. 168).

This Chapter looks at the gaze organization of mother and infant at three and six months of age. In determining whether gaze behavior is educated following questions guided the present analysis:

1. What are the mothers' and infants' patterns of gaze in early interactions?

2. Do these patterns change over time?

3. Is this change the result of the interaction?

4.2. Method

4.2.1. Participants and Data

For this analysis, the corpus consisted of the videos from the first and fourth visit at the families' homes, i.e., when the infants were three (3 months ; 12 days, *SD* = 5 days) and six months old (6 months ; 4 days, *SD* = 7 days), respectively. The first visit was selected so as to get an idea of the mother-infant gaze patterns in the interactions with the youngest infants in the sample. The fourth visit was selected as a point of comparison, in which the development in the gaze patterns (if any) would be visible. Because it has been suggested that infants' gaze at six months shifts from being interpersonal and dyadic to include objects (Bakeman & Adamson, 1984), it seemed the correct age to check for possible development in the gaze behavior of the infants and the mothers. All seventeen mothers and their infants were considered for this analysis. Out of nine boys and eight girls, twelve were firstborn (five boys and seven girls).

The average duration of the video recordings during the first visit was 449 seconds *(SD* = 233). The shortest recorded session was 100 seconds and the longest 1026 seconds. For the fourth visit, the average duration was 481 seconds. The shortest recorded session was 141 seconds and the longest 1000 seconds.

4.2.2. Data analysis and coding

As already elaborated in the previous section, data analysis was a multiple-step process: in a first step, single instances of interest were collected as qualitative observations. To visualize and document the gaze behavior in these qualitative observations, ELAN transcription software (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) was used. This enabled visualization of the sequencing of mother and infant gaze behavior. Furthermore, sequences were transcribed using a turn-by-turn transcription method inspired by GAT-2 (Selting et al., 2009) in which snapshots taken from the videos are included to visualize the coordination of the gaze behavior with the ongoing interaction as reported in the transcripts (for a similar method, see

Pitsch et al., 2014; Pitsch, Vollmer, & Muehlig, 2013; Li, 2013). These qualitative observations will be presented first.

In a second step, these observations were generalized using microanalytical continuous coding (for a detailed methodological account, see de Barbaro, Johnson, Forster, & Deak, 2013; Bakeman & Quera, 2011). For this, the entire data were coded with INTERACT®9 (http://www.mangold-international.com), an observational data transcription software. Frame-to-frame analysis was used to mark the onset and offset of gaze intervals. Gaze direction was coded for both mother and infant. Three possible gaze locations were coded:

- gaze at (other): this includes intervals in which the participant's gaze towards the other participant's face was visible.

- gaze away: this includes intervals in which the participants' gaze was directed outside the cameras' field of view.

- gaze at object: this includes intervals in which the participant gazed at parts of the other participant's body beyond the face, i.e., the hands or torso, or at objects within the cameras' recording range.

Intervals during which the eyes were blocked from the cameras' view were not coded. Subsequently, periods of overlap between *'infant gaze at'* and *'mother gaze at'* were calculated and new annotations generated automatically with the INTERACT®9 software. This yielded intervals during which mother and infant directed their gaze towards each other's face simultaneously. Such intervals were labelled *'eye contact'*.

The coded data were then further analyzed using descriptive statistics and statistical tests. This analysis included the percentage of the interaction time for each gaze location and the average duration of gaze events for each gaze location. Further statistical tests were also used to compare the data from the first and fourth visit.

4.3. Results

4.3.1. Three-month-old infants

Qualitative Microanalysis

Example 1

VP17 1st visit (00:23-00:46)

In this example, the three-month-old infant is looking at the mother for the duration of the event. The mother due to the nature of her activity of diapering the infant (wetting her cleaning cloth, wiping the infant's skin, etc.) has to switch between looking at the infant and looking away. When the mother's gaze returns to the infant, the infant's gaze has never left the mother. As illustrated in the following transcript, the mother regularly checks back on the infant, looking at her face, but then always looks down again while continuing the activity.

Transcript 1

In line 1, the sequence begins with the mother looking towards her back (image a), where she has a bowl filled with water and is wetting her cleaning cloth. She looks up to the infant (image b) while uttering her sentence and then looks down (image c) towards the infant's body, where she will next use the cleaning cloth held in her hands.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
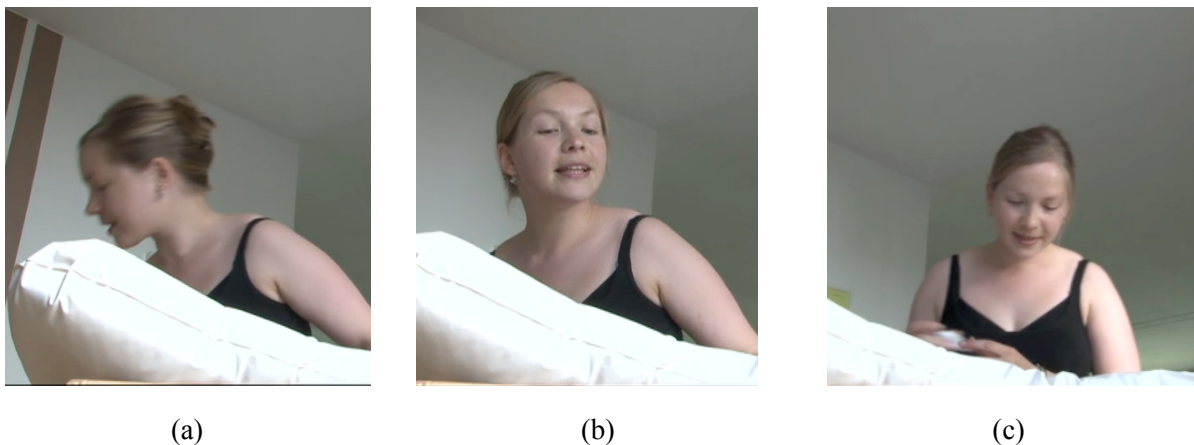
Figure 4.1.: Snapshots referring to Transcript line 1

```
1 M: *ᵃ¹das wird *ᵇsich lohnen, hä? *ᶜ(1.7 s)
```
this will be worth it, huh? (1.7 s)²

(d)   (e)
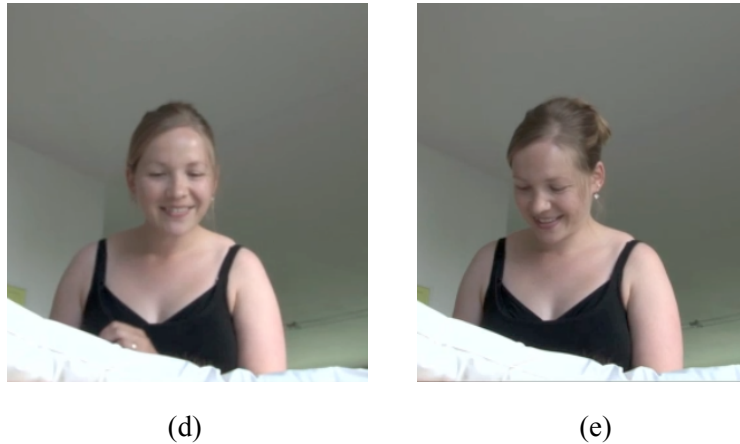
Figure 4.2.: Snapshots referring to Transcript line 2

```
2 M: *ᵈja was *ᵉ (2.9 s)
```
yes, what (2.9 s)

In line 2, the mother looks up at the infant, smiling at her before speaking (image d) and proceeds with looking at her hand, unfolding the cleaning cloth in her hand (image e).

```
3 M: ja [wa
```
yes [wha
```
4 I: [((vocalization))]
```

---

[1]The letters in superscript correspond to the image with the same letter in the figure above. This convention is used throughout the entire text.

[2]The original utterances are in German. An English translation is always provided below each utterance. The translation is a word-to-word exact translation where possible. Otherwise, the goal is to communicate the meaning of the utterance.

(f)   (g)   (h)
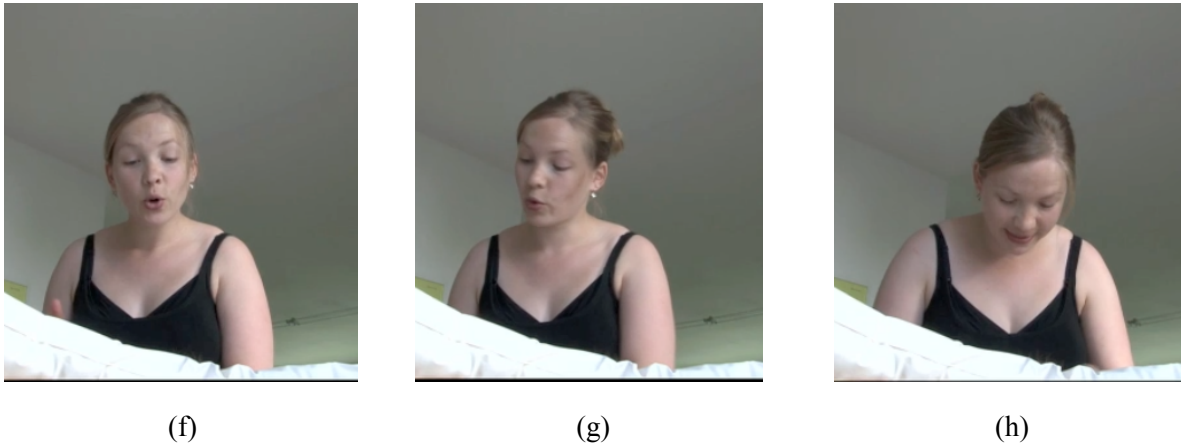
Figure 4.3.: Snapshots referring to Transcript line 5

```
5 M: *f ((sound)) *g (.)((sound))(.)guck mal *hwas jetzt kommt
  ((sound)) (.) ((sound)) (.) look what is coming now
```


(i)   (j)
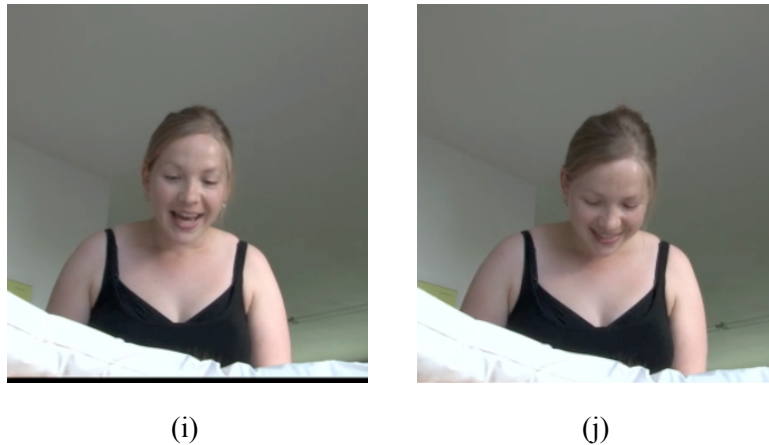
Figure 4.4.: Snapshots referring to Transcript line 6

```
6 M:((inbreathe)) *i was kommt jetzt? *j (1.4 s)
  ((inbreathe)) what is coming now? (1.4 s)
7 M: was [kommt je]tzt?
  what is coming now?
```

```
8 I: [((vocalization))]
```

In line 4, the infant vocalizes, upon which the mother looks up at her in the beginning of line 5 (image f) and responds to the infant vocalization by producing a sound similar to that of the infant. She returns to her activity while repeating the sound once more (image g) and proceeds with looking down towards the infant's body to clean her (image h). What follows is a repetitive sequence in which the mother keeps repeating the phrase *"look what is coming now"* (lines 5-7), while alternating her gaze direction between the infant's body and the infant's face; here again the mother is smiling when looking at the infant and looks away to continue her action (images h, i and j).
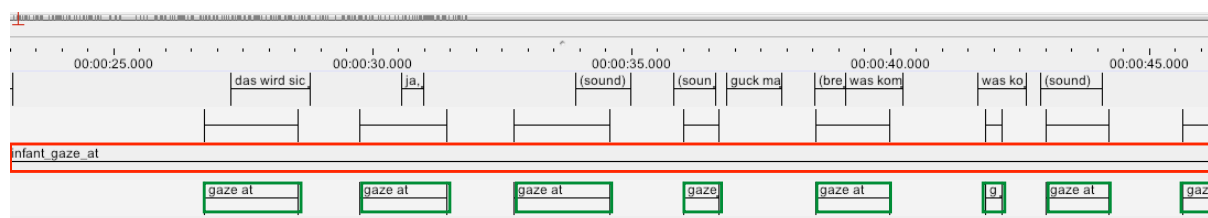
Figure 4.5.: Detail from ELAN transcript. The infant's gaze at the mother is marked in red and the mother's gaze at the infant is marked in green.

This gaze pattern of mother and child in this particular sequence is best visualized when looking at the extract of the ELAN transcript above (see Figure 4.5). In the present sequence, there is no gaze switching on behalf of the infant and only the mother switches her gaze between the infant's face, the infant's body, and objects used in the activity.

Although the above presented pattern occurred quite often in the data, not all infants spent such long intervals looking at their mothers' face, and even for those who did, they did not do so for the entire interaction. So there were also many examples of more active gaze exchanges in the interactions with the three-month-olds. Still, the observation was that the gaze of mother and infant kept returning to each other's face as will be illustrated in the following example.

Example 2

VP10 1ˢᵗ visit (00:00-00:16)

This is the opening sequence of this recording. The mother begins the activity by taking off the infant's sleeping bag. In this case, the mother's gaze follows the pattern already described above, constantly changing direction between the infant's face, the infant's body, and the objects involved in carrying out the activity. However, the infant in this case is active in also changing the direction of his gaze as the mother proceeds with the activity.

Transcript 2

(a)   (b)
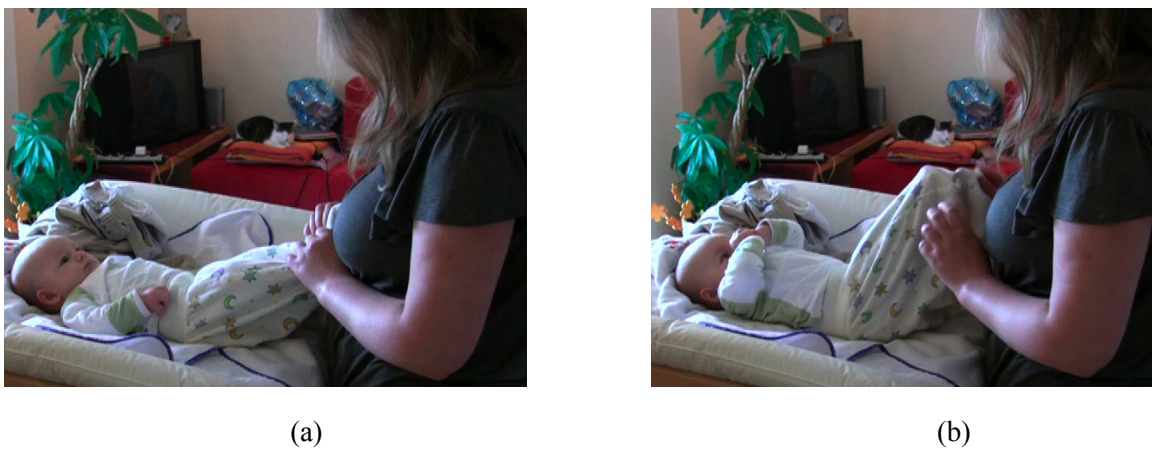
Figure 4.6.: Snapshots referring to Transcript line 1

```
1 M: ((imitation)) (0.6 s) *ªziehen wir jetzt mal den=
   =Schlafsack *ᵇaus? (1.1 s)
   ((imitation)) (0.6 s) shall we take the sleeping bag=
   =off now? (1.1 s)


2 M: he? *ᶜsollen wir den Schlafsack *ᵈausziehen? (1.5 s)
   hm? should we take off the sleeping bag? (1.5 s)
```

The sequence begins with the mother proposing to take off the infant's sleeping bag (line 1). The

(c)   (d)
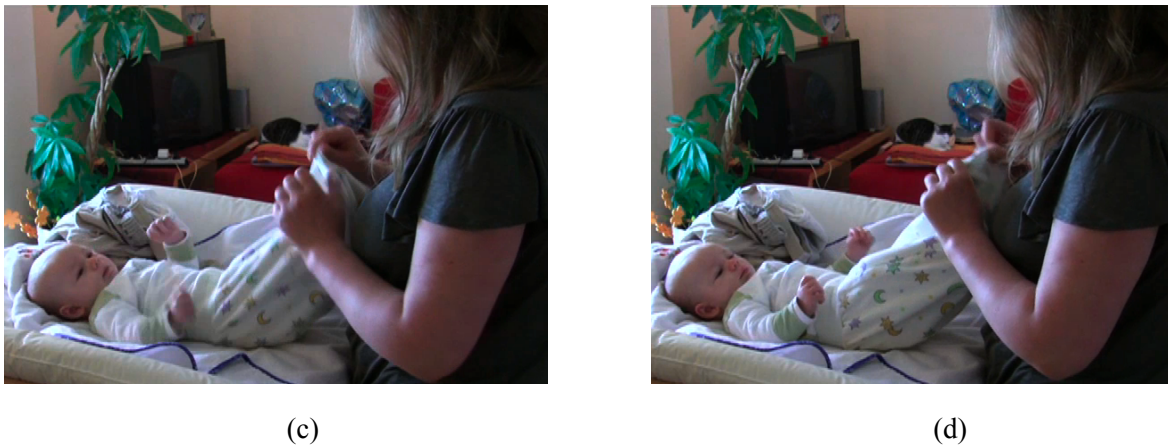
Figure 4.7.: Snapshots referring to Transcript line 2

infant is looking at the mother (image a) and then moves his arms in front of his eyes and looks up (image b) only to return his gaze to the mother's face in (image c). While the mother repeats the question about taking off the sleeping bag (line 2), she begins to move it. The infant diverts his eyes from the mother's face to her hands and looks at the sleeping bag, too (image d).

(e)
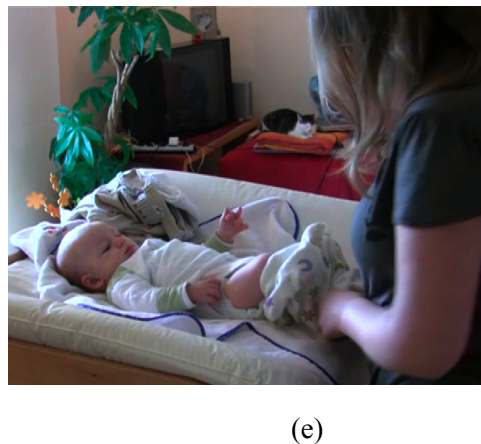
Figure 4.8.: Snapshots referring to Transcript line 3

```
3 M: ((inbreathe)) *e(2.2 s)
```

While the mother proceeds with this action (line 3), the infant monitors it, still looking towards

the mother's hands and/or the sleeping bag (image e).

(f) (g)
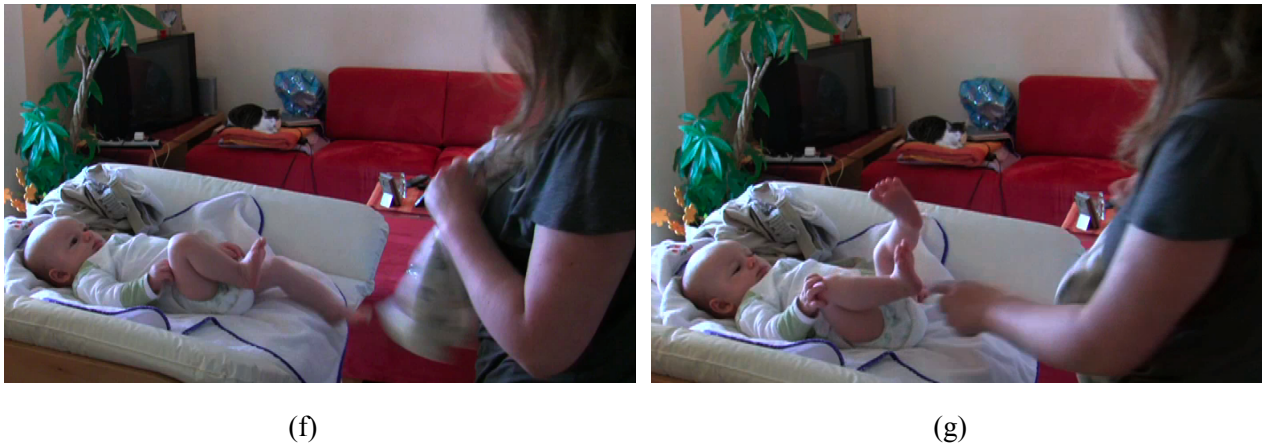
Figure 4.9.: Snapshots referring to Transcript line 4

4 M: so *f(– – –) guck mal die Oma*g hat eben vergessen den Knopf=
  so (– – –) look granny forgot to close the button before=

(h)

Figure 4.10.: Snapshot also referring to Transcript line 4

  = zu zu machen *h(.) he? (2.3 s)
  =(.) huh? (2.3 s)

Having taken off the sleeping bag, the mother steps back. At this point, the infant looks at the mother's face again (image f). He lifts his legs up and shifts his gaze from the mother's face to his legs (image g). In line 4, the mother comments on the fact that one of the buttons on the infant's onesie was left unbuttoned, probably when the infant's grandmother changed him the last time. Upon finishing her utterance, the mother again moves forward and leans with her upper body over the infant. Mother and infant engage in eye contact (image h).

The gaze pattern is also illustrated in Figure 4.11. Here it becomes visible that the gaze of the infant is distributed between gazing towards the mother's face and gazing at his body, her body, and the objects involved in the scene (in this case the sleeping bag).
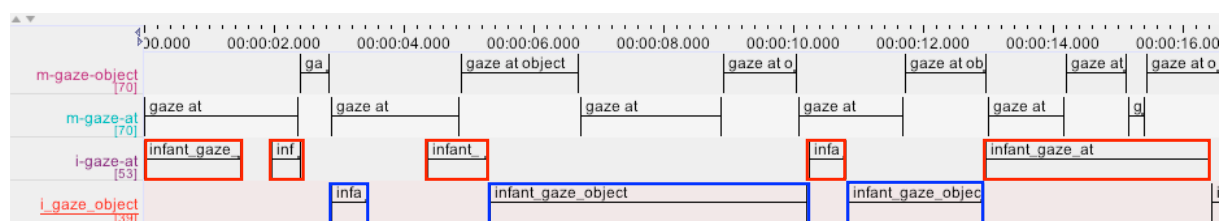
Figure 4.11.: Detail from ELAN transcript. The infant's gaze at the mother is marked in red and the infant's gaze at objects or the mother's body is marked in blue.

In this sequence, the infant is active in switching his gaze. Interestingly though, a review of the details described above shows that the gaze of the infant is nonetheless restricted to a specific region within his field of view (see Figure 4.12).

Even when he is looking at objects or the mother's body, he is changing the angle of his head mostly upwards or downwards and only slightly to the side. Thus, even when the infant is not gazing at the mother's face, his body and head orientation remain very similar to the face-to-face orientation. This was observed in the same sequence in the next transcript line, shown below. Here, the mother is holding the infant's legs and rubbing them, and the infant switches his gaze to his feet (image i). His gaze remains at his feet for the duration of lines 5 and 6.

```
5 M: und warme Füße*i hast du a::uch, ist alles in=
```
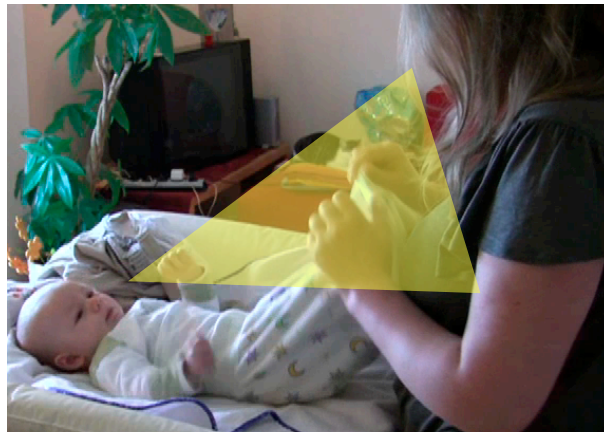
Figure 4.12.: Illustration of the area in which the gaze of the infant moves. The yellow triangle
represents this area.

(i)

Figure 4.13.: Snapshot referring to Transcript line 5

```
  =Ordnung (1.9 s)
  and warm feet you have a::lso:: everything is OK (1.9 s)
6 M: das sind deine Füße (2.8 s)
  these are your fee::t (2.8 s)
```

Next, the mother starts to search for the clean diaper she wants to put on (line 13). At this point,

her movement leaves the face-to-face orientation space as she starts picking up the clothes to the side of the infant and behind his head. This time, the infant does not switch his gaze but remains gazing at the mother's face (image j).

```
13 M: Wo habe ich jetzt die Windel hingetan? (0.6s)
```
  where have I put the diaper now? (0.6)s

(j)

Figure 4.14.: Snapshot referring to Transcript line 14

```
14 M: ((inbreathe)) da: *jliegt sie (1.1 s)
```
  ((inbreathe)) the::re it is (1.1 s)

In line 15, the mother has found the diaper and positions it within the face-to-face gaze area, as the infant switches his gaze from her face to the diaper (image k).

(k)

Figure 4.15.: Snapshot referring to Transcript line 15

```
15 M: *kja? (0.5 s) diesmal mit einer Raupe drauf? (0.6 s)=
      =ist OK?
      yes? (0.5 s) this time with a caterpillar on it? (0.6 s)=
      =OK?
```

Quantitative Analysis

The coding scheme for interactions between the mothers and their three-month-old infants included the following descriptive statistics:

1. Percentage of the interaction time for each gaze location (for mother and infant)

2. Average duration of gaze events for each gaze location (for mother and infant)

3. Percentage of the interaction time spent in eye contact

4. Average duration of eye contact events

The percentage of interaction time illustrates the extent to which mothers and infants looked at the various gaze locations, whereas the average duration of gaze events provides an impression

of the pace of gaze organization.

The mother gazed at the infant's face most (52.91% of the time ; *SD* = 16.33). This behavior was followed by gazing at objects or the infant's body (40.99% of the time ; *SD* = 17.14). This confirms the qualitative observations presented in the previous section in which the mother switched her gaze constantly between the infant's face, his or her body, and the objects involved in the activity. Finally, mothers spent very little time gazing away (2.16% of the time ; *SD* = 2.35).

The three-month-old infants also gazed mostly at the mother's face (58.29% of the time ; *SD* = 20.90). Here, the maximum value reached about 94 % of the interaction time. Gaze at objects took up a much smaller proportion of the interaction (18.35% of the time ; *SD* = 19.10). Finally, infants gazed away even less than that (12.89% of the time ; *SD* = 11.74).

The average duration of gaze events for the mother resembles the above results. The mothers' duration of gaze events toward the infant was the longest (3.62 sec. ; *SD* = 2.03). This was followed by gaze durations at objects (2.71 sec. ; *SD* = 1.37) and then intervals in which mothers gazed away (1.35 sec. ; *SD* = 0.94).

For the infants, the average duration of gaze events shows a similar pattern: gazes at the mother were the longest (10.26 seconds ; *SD* = 8.05), followed by the durations of gaze at objects (4.13 seconds ; *SD* = 4.15), with the gaze intervals away being the shortest (3.44 seconds ; *SD* = 3.03).

The complete descriptives can be found in the Appendix Table A.1.

For the three-month-olds, eye contact occurred about one third of the interaction time (34.84% of the time ; *SD* = 14.03 ; *MIN* = 8.61 ; *MAX* = 66.06). These eye contact events lasted on average about 3 seconds (*SD* = 1.79 ; *MIN* = 0.97 ; *MAX* = 7.08). Again, the complete descriptives can be found in Appendix Table A.3.

4.3.2. Six-month-old infants

Qualitative Microanalysis

Example 3

VP17 4th visit (00:00-00:40)

Transcript 3

(a) (b)

Figure 4.16.: Snapshots referring to Transcript line 1

```
1 M: *a hm? *b (.) na das kennst du in zwischen hä? (3.2 s)
   you know this in the meantime, huh? (3.2 s)
```

This example is from the same mother-infant dyad as example 1 after three months (when the baby was six months old). In this example, there are differences in the gaze patterns. Here, the infant is more active in quickly switching her gaze to and away from the mother. There are more gaze intervals and they are shorter.

The sequence starts with the infant gazing at the mother (line 1, image a). However, her gaze does not stay with the mother as she begins to utter her phrase, commenting that the infant knows the process by know, it being the 4th visit at the family's home in which the infant's diaper was changed in this particular location. The infant at once switches her gaze to the side and looks

away behind the changing table (image b), exploring the room from her perspective. In Line 2, the same pattern is repeated. The infant's gaze returns to the mother, but only very shortly as she very quickly looks away again, this time to the other side of the room (images c and d). As soon as the mother begins her utterance, the infant shifts her gaze to the other side.

(c)   (d)

Figure 4.17.: Snapshots referring to Transcript line 2

```
2 M: *cdas kennst *ddu inzwischen (2.6 s)
  you know this in the meantime (2.6 s)
3 M: hm? (1.5 s) was siehst Du denn da? (4.5 s)
  hm? (1.5 s) what do you see there?
```

In line 3, the mother comments on the infant's very active gaze away from the changing table, asking her what it is she is looking at. In line 4, the infant looks again shortly at the mother (image e), who grasps the opportunity to make eye contact with the infant by leaning a bit forward and tilting her head to the same angle as the infant's head (image f) establishing a straight line in the face-to-face orientation between herself and her infant and marking this by saying *"Kuck kuck"*, which is the phrase used in German when playing peek-a-boo. Right after this, the infant's gaze drifts away again (image g).
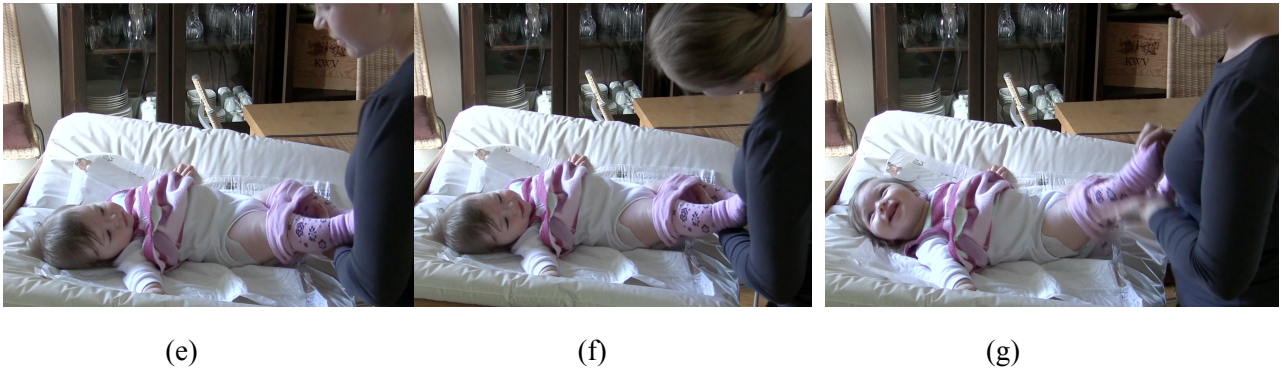
Figure 4.18.: Snapshots referring to Transcript line 4

```
4 M: *ekuck kuck *f, *g(4.5 s)

5 M: erstmal rumgucken überall, hä? (- -)
   first look all over, huh? (- -)
```


Figure 4.19.: Snapshots referring to Transcript line 6

```
6 M: *h((inbreathe)) ja machen wir *imal den Popo frei (6.4 s)
   ((inbreathe)) yes, let's make your bottom free (6.4 s)
```

The same sequence occurs in line 6 (images h and i). It is only in line 7 that the mother succeeds

in capturing the infant's visual attention beyond a fleeting response, by tickling her lightly on the belly (images j and k) while calling her by repeatedly saying *"hey"*.

(j)   (k)

Figure 4.20.: Snapshots referring to Transcript line 7

```
7 M: *ʲE:Y (- -) *ᵏey (.) ey (.) ha ha ha Zeig mal
     he:y (- -) hey (.) hey (.) ha ha ha show me
```

The gaze pattern between mother and infant in this example is best illustrated in Figure 4.21. Here, it becomes apparent that the infant is looking at the mother's face only for short intervals and some of the mother's gazes at the infant also go completely unnoticed by the infant.

Figure 4.21.: Detail from ELAN transcript. MG = maternal gaze; IG = infant gaze. Marked in red is the infant's gaze at the mother and marked in green the mother's gaze at the infant

Also in this example, the infant's changing interest becomes apparent. The infant looks back

at the mother's face and smiles upon making eye contact with her or upon hearing an utterance (images b, g, i, j, k), but does not engage in prolonged episodes of eye contact. Instead, she seems more interested in exploring her environment, in this case in looking at the surroundings. Another change in the infant's gaze is the exploration of the objects involved in the changing activity, shown in the next example.

Example 4

This sequence begins with the mother dressing the infant. In particular, she has turned him to the side to facilitate putting on his tights (image a). In this position, the infant grasps the box of wet wipes lying on the changing table. He flips back on his back, looking at his mother and at the box, and continues to touch and manipulate the box (images b and c). The mother responds to this interest by moving the box within the infant's reach. She first places the box on his belly for him to play with it, which doesn't prove to be very easy for him, perhaps because the box is too large for him to handle.

VP11 4th visit (05:22-06:30)

Transcript 4

(a)   (b)

Figure 4.22.: Snapshots referring to Transcript line 1

```
1 M: *ª[((inbreathe))*ᵇ(0.3 s) nein? (0.4 s)
   [((inbreathe)) (0.3 s) no? (0.4 s)
2 I: [((infant breathing heavily))
```

(c)

Figure 4.23.: Snapshot referring to Transcript line 3


```
3 M: kommst du wieder zurück?*c (0.3 s) ((imitation))
     are you coming back? (0.3 s) ((imitation))
```


This is followed by a long pause in the mother's speech in which she adjusts the infant's position on the changing table.

(d)

(e)

(f)

Figure 4.24.: Snapshots referring to Transcript line 5


```
5 M: *d((sound)) (3.3 s)*e da ist die Kiste*f (1s) da ist sie
     ((sound)) (3.3 s) there is the box (1 s) there it is
```

In line 5, The mother has leaned forward, approaching the infant in a face-to-face orientation. The infant is still touching the box of wet wipes, but cannot look towards it (image d). The mother takes hold of the box and positions it on the side (images e and f). At the same time, she comments on her action by naming the object of the infant's interest, namely, the box of wet wipes. What follows is a prolonged episode of the infant gazing uninterruptedly at the box and touching it.

(g)

Figure 4.25.: Snapshots referring to Transcript line 8

```
8 M: schön*ᵍ bunt (1 s) grün (0.3 s) blau (0.5 s) lila ist=
  =unten drunter (0.6 s)
  nice and colorful (1 s) green (0.3 s) blue (0.5 s) purple=
  =is below (0.6 s)
```

In line 8, the mother turns the box around so that the inscription on the box is made visible to the infant (image g), then starts to describe what she sees on the box, acknowledging the infant's interest in the object and providing him with details on its appearance and features (lines 8 - 12). She also accompanies these descriptions by using pointing gestures. In line 9, she points at the purple color purple (image h), then the turquoise (image i).

(h)   (i)

Figure 4.26.: Snapshots referring to Transcript line 9

```
9 M: ((inbreathe)) (.) da ist*ʰlila (- -) da ist ro::t (- -)=
  =da ist gelb und*ⁱda ist türkis(- - -)
  ((inbreathe)) (.) purple is there (—) red is there(- -)=
  =yellow is there and turquoise is there (- - -)
10 M: ((inbreathe)) (0.4 s) hm (1 s) oder auch grün (3.1 s)
  ((inbreathe)) (0.4 s) hm (1 s) or maybe green (3.1 s)
```

Next, with the infant still focused on the object, she flips the box again to reveal another side and starts to describe it (lines 11 and 12), providing more information about the infant's object of interest. In line 12, she points to the picture of the baby on the package (image j), and then lets the infant manipulate the object himself as she provides the deictic expression *"there"* (image k).

```
11 M: hast du gesehen dass da Babies drauf sind? (1.6 s)
  have you seen that there are babies on it? (1.6 s)
```

(j)   (k)

Figure 4.27.: Snapshots referring to Transcript line 12

```
12 M: guck*ʲmal da, ein ganz kleines Baby (2.2 s) ein ganz=
   =kleines Bild von einem Baby (3 s) da ist noch=
   =ein Baby (- -) da*ᵏ
    look there, a very small baby (2.2 s) a very small photo=
   =of a baby (3 s) there is also a baby too=
   =(- -) there
```

In this example, we see a different gaze pattern. It resembles the gaze behavior of the three-month-old infant presented above, only now this gaze pattern is observed when looking at objects and not at the mother's face. This is best illustrated in the detail of the ELAN transcript in Figure 4.28, in which the infant's prolonged gaze at the object is clearly visible (marked in blue), whereas the mother keeps shifting between gazing at the object (marked in orange) and checking back at the infant's face (marked in green).

The examples presented are indicative of a change in the gaze patterns of both mother and infant. It has been shown that at the age of six months, the durations of gaze intervals are shorter and the infants seem to become increasingly interested in the surroundings and objects involved in the interaction. In the next section, the quantitative analysis of gaze behavior for the 4th visit will be presented.
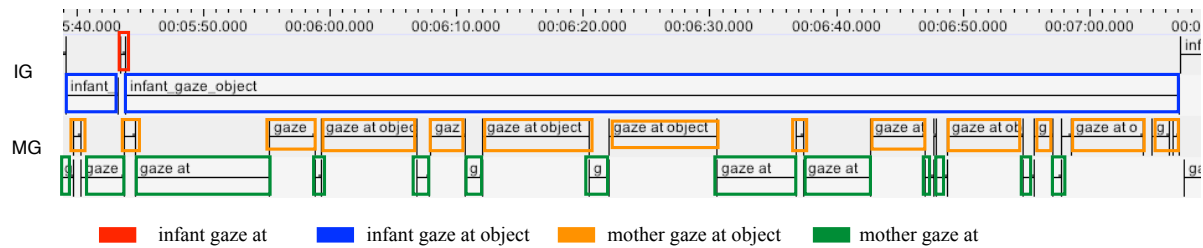
Figure 4.28.: Detail from ELAN transcript. MG = maternal gaze; IG = infant gaze. The gaze behavior of mother and infant are marked in different colors.

Quantitative analysis

The analysis of this data followed the same process already described above in the section describing the interactions with the three-month-old infants.

The mother gazed at the infant's face most (46.96% of the time ; $SD = 20.26$). Gaze at objects occurred slightly less often (45.71% of the time ; $SD = 21.33$). Finally, gaze away was scarce (1.91% of the time ; $SD = 1.66$).

For the infants, gaze at the mother was the most used gaze location (32.89% of the time ; $SD = 16.66$). Infants looked slightly less at objects (31.00% ; $SD = 17.78$), and gazed away a little less (22.66 % of the time ; $SD = 10.00$).

When looking at the duration of gaze events of the mothers, these were equally long for gazing at the infants' face and looking at objects (2.62 seconds ; $SD = 2.24$ and 2.62 seconds ; $SD = 1.83$, respectively). Intervals in which mothers gazed away were shorter (1.35 seconds ; $SD = 0.92$).

For the infants the longest intervals were found for infants' gaze at objects (4.24 seconds ; $SD = 3.73$). These were followed by the intervals during which infants gazed away (3.62 seconds ; $SD = 1.69$), while gazes at the mother were only slightly shorter (3.61 seconds ; $SD = 1.65$).

The complete descriptive data can be found in the Appendix Table A.2.

Furthermore, in the interactions with the six-month-old infants eye contact occurred about

one fifth of the interaction time (20.53% of the time ; SD = 13.66) with quite short average duration of eye contact events (1.80 seconds ; SD = 0.76).

The complete descriptive data can be found in the Appendix Table A.4.

4.3.3. Comparison three- vs. six-month-old infants

This section includes the comparative quantitative analysis of the gaze behavior of mother and infant at three and six months. The qualitative and quantitative results will be discussed in a later section.
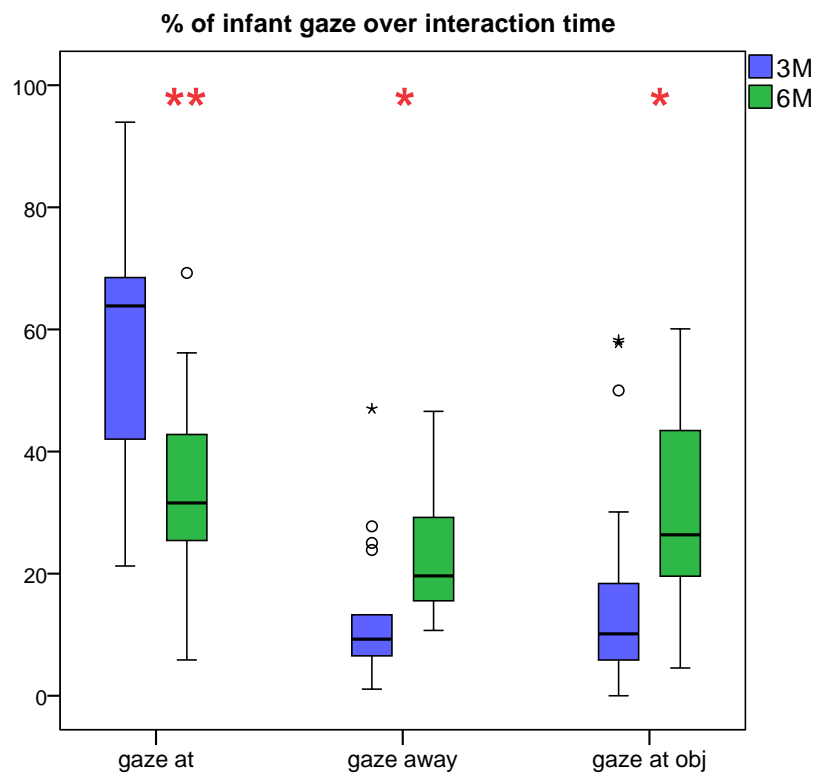
Figure 4.29.: Infant gaze at three and six months of infants' age. The red asterisk denotes the significant differences.

As illustrated in Figure 4.29 the six-month-old infants spent significantly less time looking at their mother's face: for INFANT GAZE AT *t(16)= 4.118 ; p = .001*. They spent significantly more time looking at their mother's hands, her body, objects (INFANT GAZE AT OBJECT *t(16)= 2.432 ; p = .027*), and the surroundings (INFANT GAZE AWAY *t(16)= 2.319 ; p = .034*) The complete data set can be found in Appendix Table A.5.

Average duration of infant gaze events (in sec.)
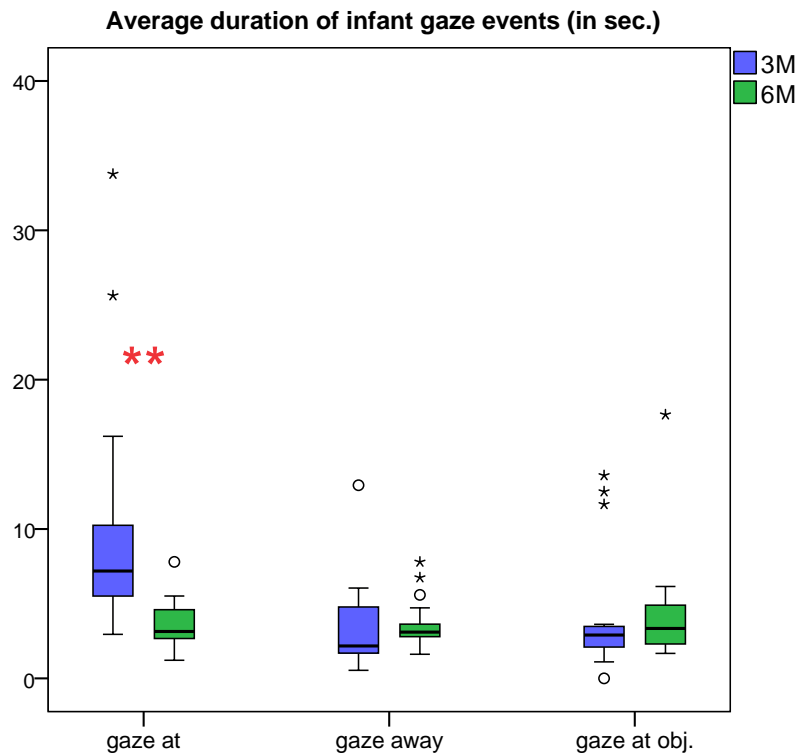
Figure 4.30.: Average duration of infant gaze events at three and six months of infants' age. The red asterisk denotes the significant differences.

In Figure 4.30, it also becomes apparent that the duration of infants' gaze events at the mother grew significantly shorter: *t(16)= 3.414 ; p = .004*. The complete data set can be found in Appendix Table A.6.

The mothers' behavior when their infants were six months old resembles that of the infant, i.e., less gaze at the other's face (her infant's face) and more gaze at objects. Nevertheless, the

differences did not attain significance (see Figure 4.31).

This indicates that the mothers did not modify their gazing behavior overall significantly. With respect to mothers' average durations of gaze events, there was a significant decrease of event durations for gaze at the infant: $t(16)$=2.947 ; p = .009 (see Figure 4.32).
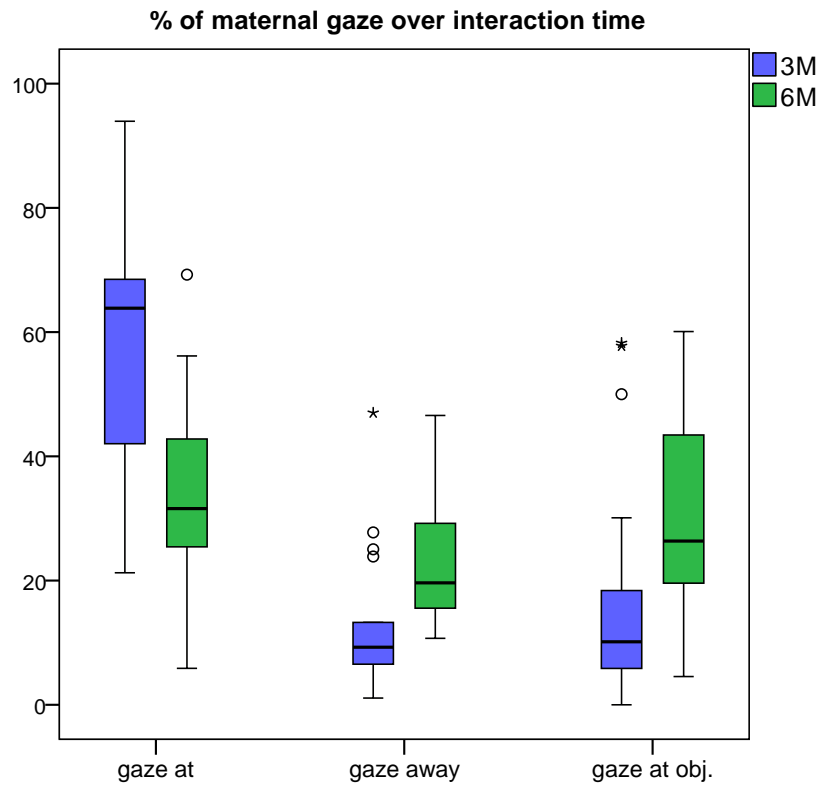
Figure 4.31.: Maternal gaze at three and six months of infants' age.
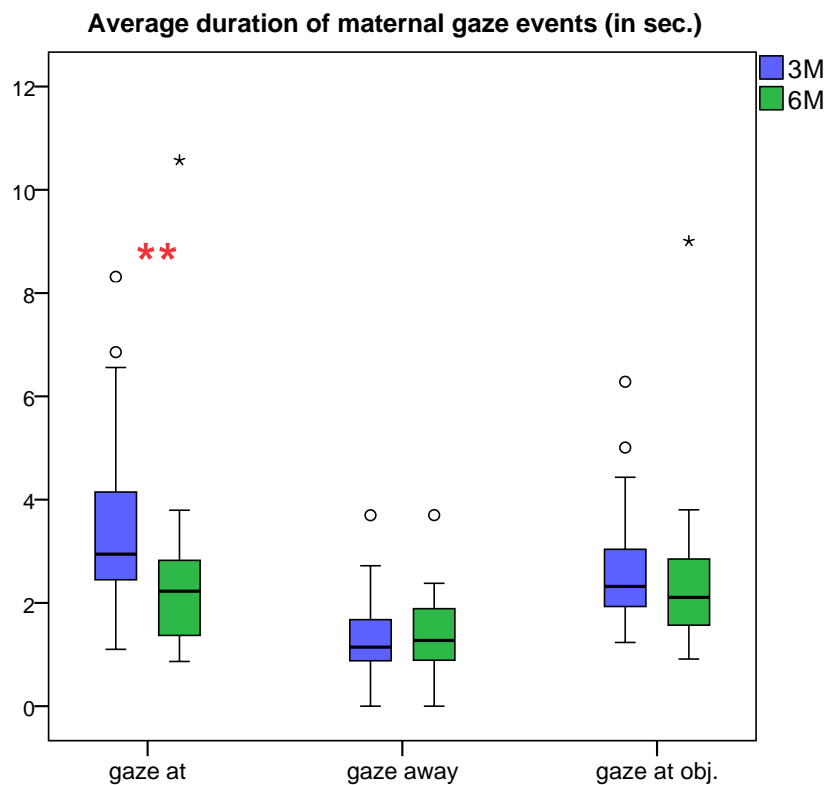
Figure 4.32.: Average duration of maternal gaze events at three and six months of infants' age. The red asterisk denotes the significant differences.

At six months, mother and infant looked at each other significantly less than when the infant was three months old: For EYE CONTACT: $t(16)=4.152$; $p = .001$. Also, the average duration of eye contact intervals became significantly shorter in the interactions with the six-month-olds : $t(16)=3.814$; $p = .002$ (see Figures 4.33a and 4.33b, respectively).
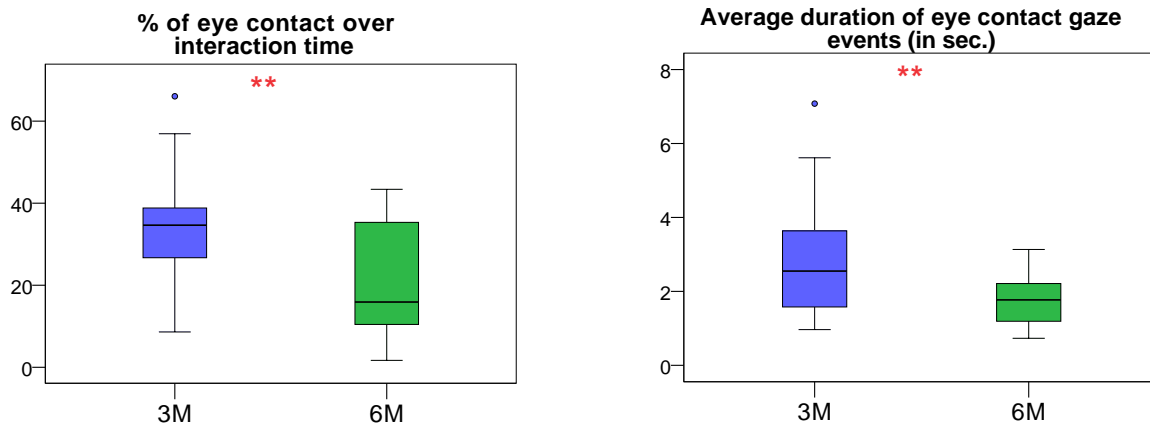
(a) Proportion of eye contact over interaction time

(b) Average duration of eye contact events

Figure 4.33.: Eye contact at three and six moths of infants' age. The red asterisk denotes the significant differences.

To find out whether gaze patterns change or remain constant over time and whether there was continuity in the behavior of the mothers and the infants, the data from the interactions when the infants were three and six months old were correlated.

Table 4.1 presents the paired correlations for the three-month-old and six-month-old data. The mothers show a continuity in their behavior; the three-month-old data correlate significantly with the six-month-old data for percentage of gaze at the infant ($r = .647$; $df = 16$; $p = .005$), and average duration of gaze at the infant ($r = .787$; $df = 16$; $p < .000$). For mothers' gaze at object the data correlate significantly for percentage of gaze at objects ($r = .700$; $df = 16$; $p = .002$), but not for average duration of gaze at objects. The data for gaze away did not correlate significantly.

For the infants, appears to be a 'break', or discontinuity, between the two visits as the infants who behaved in a particular way at three months, behave differently at six months. None of the gaze location data at three and six months correlate significantly.

Finally, for eye contact, the data also correlate significantly for average duration of eye contact events ($r = .761$; $df = 16$; $p < .000$) and are marginally significant for eye contact $r = .474$; df

Table 4.1.: Paired Sample Correlations for GAZE at three and six months

Variables Pairs		N	Correlation	Sig.
M-gaze at 3M and 6M	(%)	17	**.647**	**.005**
	(Avg. in sec.)	17	**.787**	**.000**
I-gaze at 3M and 6M	(%)	17	.097	.711
	(Avg. in sec.)	17	.118	.651
M-gaze at object 3M and 6M	(%)	17	**.700**	**.002**
	(Avg. in sec.)	17	.280	.277
I-gaze at object 3M and 6M	(%)	17	.324	.204
	(Avg. in sec.)	17	-.157	.546
M-gaze away 3M and 6M	(%)	17	.377	.136
	(Avg.)	17	.356	.161
I-gaze away 3M and 6M	(%)	17	-.272	.291
	(Avg. in sec.)	17	.154	.556
Eye contact 3M and 6M	(%)	17	**.474**	**.055**
	(Avg. in sec.)	17	**.761**	**<.000**

= 16 ; *p* = .055).

To determine whether there was already a relationship between mothers' and infants' gaze towards each other, the data was also cross-correlated, i.e., mothers' gaze at the infant was correlated with infants' gaze at the mother.

At six months, there is a significant correlation between the percentage of mothers' gaze at their infants and the infants' gaze at the mothers (*r* = .485 ; *df* = 16 ; *p* = .049). This means that mothers who gazed more at their infants had infants who gazed more at them. For the three-month-olds this correlation is not significant. This suggests that mothers and infants might actually affect each other's attention as their gaze patterns converge over time, as a result of their history of interaction. The way in which this may happen motivated the gaze analysis presented below.

4.3.4. Framing attention: mother-infant gaze roles

Motivated by the finding that mothers and infants might affect each other's attention, the next step was to focus on the adaptiveness of gaze patterns. One aspect of the interaction which became evident through the analyses presented above was that although the mother did not significantly modify the overall proportion of the interaction she spent looking at the infant, she did modify the duration of gaze intervals, thus aligning with the faster organization of gaze taking place in the interactions with the six-month-olds.

One striking observation was that a comparison of three- and six-month-old infants seemed to reveal a role reversal concerning who is the one maintaining a more constant gaze and who is the one moving in and out of eye contact. The observation was that for the three-month-olds, it was the infant who was looking at the mother's face for long periods, whereas it was the mother who set the pace for the organization of eye contact: She was the one who left the mutual gaze episodes to continue with her activity of diapering the infant, that is, wetting her cleaning cloth and cleaning the infant's skin. When she looked back at the infant, the infant was already looking at her. At six months of age, a different pattern emerged in the mother-□infant

dyads. The infants no longer framed the gaze episodes. Rather, they engaged more actively in quickly switching their gaze to and away from the mother.

To test this hypothesis over the entire sample, the intervals of eye contact were compared to identify the mothers' and infants' roles in establishing and breaking it. More specifically, all overlapping intervals between mothers' gazes at infants and infants' gazes at mothers were automatically identified. By comparing the starting points and endpoints of each interval for each eye contact event, it became possible to calculate, whether it was the mother or the infant who initiated and terminated the interval. Moreover, by comparing these endpoints, the type of temporal relation (see, for interval relations, Nebel & Bürckert, 1995) that these overlapping gaze intervals have with respect to one another could be calculated. These measures may make clearer whether, the mother is framing the gaze of the infant or the other way round. If it is the mother who is framing the infant's gaze, then there should be more intervals of the type 'contains' (see, Figure 4.34, and Nebel & Bürckert, 1995), in which the gaze of the mother encloses the gaze of the infant. Otherwise, there should be more of the type 'during' (see, Figure 4.34, and Nebel & Bürckert, 1995), in which the gaze of the infant bookends that of the mother.

A paired-sample *t*-test was conducted and, in line with the hypothesis, mothers ended eye contact more frequently when the infants were three months old (80% of intervals) than when they were six months old (57.61% of intervals) $t(16) = 3.08$, $p < .01$. In contrast, infants terminated eye contact more frequently when they were six months old (42.40% of intervals) than when they were three months old (20% of intervals) $t(16) = 3.08$, $p < .01$.

In the further analysis, a comparison was made between those temporal relations that indicate whether the mothers or the infants frame the gaze of the other. These intervals are graphically illustrated in Figure 4.34.

For the intervals in which the mother's gaze at the infant overlapped with the infant's gaze at the mother, the frequency of CONTAINS and DURING intervals were compared at three and six months of infants' age. For the three-month-olds, there were more DURING intervals (78.6% of intervals) and fewer CONTAINS intervals (21%) than expected by chance. Respectively, for

contains   **during**

| mother gaze |

| mother gaze |

| infant gaze |

| infant gaze |

Figure 4.34.: Temporal relations between mother's and infant's gaze. On the left interval type CONTAINS and on the right interval type DURING are illustrated, respectively.

the six-month-olds, there were more CONTAINS intervals (41%) and fewer DURING intervals (59%) than expected by chance. This shows that there seems to be a relationship between AGE and the frequency of occurrence of CONTAINS and DURING temporal relations. The corresponding chi-square test for independence revealed a marginal significance, χ^2 (1, N = 81) = 3.64, p = .056.

4.4. Summary of findings

4.4.1. Findings from qualitative analysis

The cases presented in the qualitative analysis suggest that at three months, infants show a clear preference for face-to-face interaction. It was shown that although the mother switched her gaze to and from the infant, the infant's gaze remained constantly directed at the mother's face. Furthermore, even in cases of more active gaze exchanges, the gaze of mother and infant kept returning to each other's face. More specifically, it was shown that the gaze of the infant was restricted to a specific region within his or her field of view, such that even when the infant

was not gazing at the mother's face, his body and head orientation remained very similar to the face-to-face orientation.

The examples presented from the interactions with the six-month-olds showed differences in the gaze patterns. It was shown, that the infant became more active in switching his or her gaze to, and away from, the mother. Also, it was shown that the infants looked at the mother's face only for short intervals, and did not engage in prolonged episodes of eye contact. Instead, they spent much more time looking at objects. They seemed to become more interested in exploring their environment, in this case in looking at the surroundings or at the objects involved in the diaper changing activity.

A comparison of three- and six-month-old infants revealed a role reversal concerning the gaze organization. In the interactions with the three-month-olds, it was the mother who set the pace for the organization of eye contact. At six months of age, the infants no longer framed the gaze episodes. Rather, they engaged more actively in quickly switching their gaze to and away from the mother, and, thus, they set the pace for the organization of gaze.

4.4.2. Findings from quantitative analysis

The quantitative analysis showed that mothers' and infants' behavior follow certain patterns. Both mothers and infants spent most of their interactive time looking at each others' face. They also spent a considerable amount of the interaction in eye contact. The rest of the time, they looked at each other's body and at objects involved in the interaction, with the mothers looking at objects more and their infants less. Mothers stayed focused on the interaction and rarely looked away. Infants spent more time looking away from the changing table and around the room. The patterns of gaze behavior for mothers and infants during their interactions changed over time. The overall duration of looking at the partner's face decreased with infants' age. This was statistically significant only for the infants. Moreover, at the age of six months the duration of gaze intervals became significantly shorter for both mother and infant. Also, there was an increasing interest in objects. At six months infants—but not mothers—gazed significantly

more at objects and away, but the duration of gaze intervals did change significantly.

The correlation of the gaze data revealed that mothers did not change their gaze behavior over time. This was not the case for the infants, though. The cross-correlation of the data revealed that, although the proportion of the interaction time which infants spent looking at their mother at three months did not correlate with their mothers' gaze at them, at six months their gaze behavior correlates significantly, suggesting a convergence of the behavior of mother and infant within the course of development.

Finally, the analysis of the role of mother and infant in initiating and terminating eye contact and of the temporal 'framing' relations, confirmed the qualitative observation, that mother and infant switch their gazing roles between the ages of three and six months.

The following chapter will explore the ways in which gaze could be educated within these early interactions. The focus will be on the reinforcing power of maternal input in the search for systematic practices that may frame episodes of eye contact for infants.

5. Educating attention within multimodal interactions

5.1. Objective

In the previous chapter, it was shown how the gaze patterns of mother and infant develop within interaction and are dynamically adapted as a result of their history of interaction. It was suggested that mothers and infants might actually affect each other's attention; to be specific, the gaze at each other. The goal of this chapter is to take a closer exploratory look at the interactions of the mothers and infants to discover potential maternal practices or strategies which may systematically reinforce eye contact and thus frame episodes of eye contact for infants. The expectation is that if the sensory information co-occurring with eye contact differs from that in a state of non eye contact, then it could be associated with a salient event (Nagai & Rohlfing, 2009) and receive a reinforcing value for the infant's memory and attention. Following the idea that infants' attention is socialized in everyday interactions, this chapter aims to explore the following research questions:

1. Is there evidence of education of attention?

2. Is eye contact selectively reinforced by behavioral modification in the input?

3. How does this education of attention develop as the infant develops?

5.2. Method

5.2.1. Participants and Data

For this analysis, the data originated from the same corpus described in the previous chapter, i.e., it consists of the videos from the first and fourth visit at the families' homes when the infants were three and six months old, respectively. Again, all seventeen mothers and their infants were considered for this analysis.

The methodological approach used in this chapter is exploratory and qualitative. It enables the analysis of the mother's behavior as constructed within the interactive process (Goodwin, 1979). To identify the ways in which the mother's behavior is coupled to the behavior of the infant, the mother-infant dyad has to be used as the basic unit of analysis. By closely observing the reciprocation of actions between mother and infant as they unfold, one after the other within the interaction, it is possible to understand which infant behavior elicits a certain behavior from the mother as a response (see, on sequencing, e.g., Psathas, 1995; Schegloff, 2007). This approach allows zooming into the data to observe behavior as it occurs within a concrete situation or "contextual configuration" (Goodwin, 2000, p. 1490). Thus, single instances of phenomena are collected.

5.2.2. Data analysis and coding

For this analysis, the coded data described in the previous chapter were used, i.e., the gaze data and the generated 'eye contact' intervals for the three-month-old and the six- month-old infants. Next, the coded intervals were imported into the ELAN (Wittenburg et al., 2006) transcription software for visualization and exploration. For the analysis, the episodes of eye contact between mother and infant were used as a reference point from which to launch the fine-grained analyses. The moments in which eye contact between mother and infant is established and then lost were examined. This enables the observation of potential maternal strategies in coordinating their behavior with the behavior of the infant. Here, the focus was on the verbal and nonverbal

behavior of both mother and infant. Here again, selected sequences were transcribed using the method described in the previous chapter.

In the following sections, the analysis for both age groups is presented separately. First, examples are provided of episodes in which mothers attempt to (a) maintain infant attention, (b) direct attention toward themselves, and (c) pursue eye contact. Then, a qualitative comparison of the two time points of interaction when the infants were three and six months old will be made in order to gather evidence for the dynamic adaptive character of the education of attention.

5.3. Sequence analysis and discussion

The initial observation was that during eye contact, the mothers provided a range of specific behavioral practices that select the infant exclusively as the recipient, cueing her or his attention. These include both verbal and nonverbal practices. A good example for this specific and effective practice is a greeting accompanying eye contact. Greetings are interesting, because it has been suggested that they have ostensive power in that they explicitly select the recipient of communication. In fact, recent studies have shown that infants can follow a person's change of gaze direction if they hear an infant-directed greeting beforehand (Senju & Csibra, 2008). Furthermore, Fogel (1992) characterizes greetings and openings as actions responsible for the emergence of new "perceptual texture" (Fogel, 1992, p. 515) because they become attractors for ritualized behaviors (see, also Fogel, 2006).

Example 5

VP16 1ˢᵗ visit (01:25-01:31)

In this sequence, the infant is looking at the mother for the full duration, and the mother is entering and exiting the infant's line of gaze. This is due to her activity of putting the diaper on; she alternates her gaze between the diaper and the infant. The sequence begins with one episode of eye contact in silence. Here, the mother looks up to the infant and notices that he is looking at her. This is followed by a second interval of eye contact that overlaps with the mother

greeting the infant by saying *"hello"*. Eye contact is then lost again and there is silence during this period. When resuming eye contact, the mother continues her greeting by asking *"na?"*. In German, this is a typical practice to elicit further talk: Either the speaker continues with a further question or the interlocutor is expected to say how he or she is doing (see, Schegloff, 2007). In this sequence, the mother continues by asking the infant if he is feeling good.
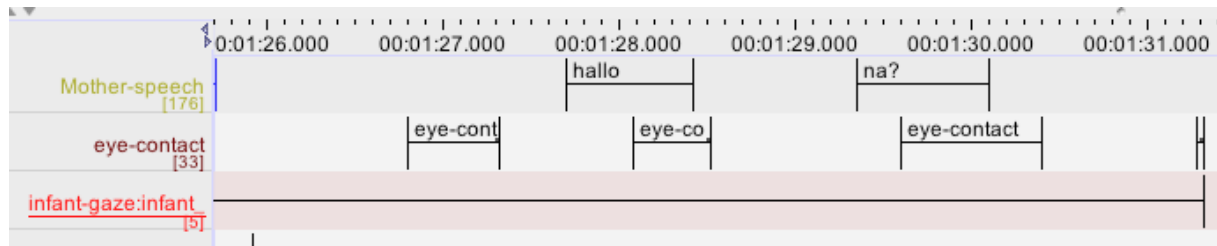
Figure 5.1.: Detail from the ELAN transcription illustrating the temporal overlap between mother speech and eye contact. The top tier includes the mother's verbal behavior and the middle tier the intervals of eye contact.

Here, the mother evidently treats eye contact as an opening of a dialogue sequence and addresses the infant by greeting him. The second time their eyes meet, she asks the infant a question, using a specific speech intonation contour explicitly selecting the infant as the addressee of her speech and handing over the turn to him. Thus, the infant experiences eye contact as overlapping with a very common everyday dialogue format—mostly including greeting and question routines—that is made more salient by accompanying nonverbal cues, as will be shown in the next example.

5.3.1. Maintaining Attention

Example 6

VP03 1st visit (00:25-00:40)

In this sequence, the mother is undressing the infant (see Figure 5.2). During the first eye contact interval, the mother uses facial expressions and head movements. During the second,

she uses facial expressions and adds a further modality—namely her hands—because she moves the infant's body in a playful manner.
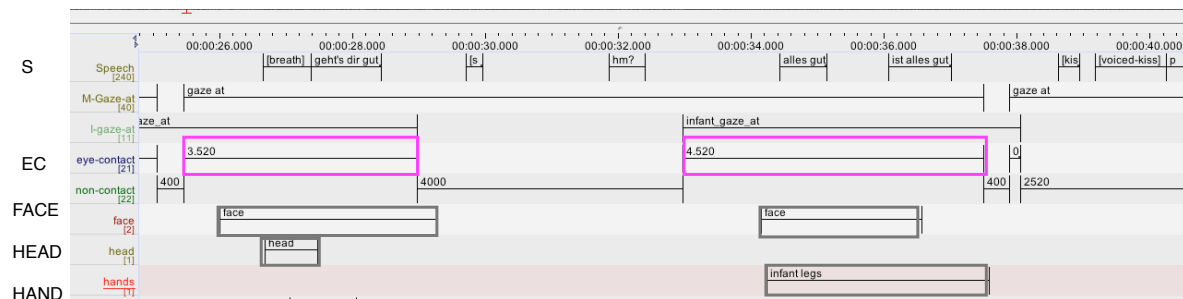
Figure 5.2.: Detail from the ELAN transcription illustrating the temporal overlap of speech (S), eye contact (EC), and movements of the mother's face, head, and hands. Eye contact intervals are marked in magenta and intervals of body movement in grey

The opening response (transcript 5, line 1), which overlaps with facial expressions and head movements, has been acknowledged in the work of M. Papoušek and Papoušek (1990). The authors underline the significance of greetings as being a simplified, prototypical behavioral pattern that is repeated often, and thus distinctive and recognizable. Furthermore, it has been suggested that tactile stimulation such as caressing the infant's legs (line 5) maintains a connection between mother and infant (Stack & Jean, 2011). In her work, Feldman (2011) suggests that "touch synchrony" (Feldman, 2011, p. 307)—the coordination of touch with gaze—helps to orient the infant toward the social environment. Also, body movement, even if induced and not self-initiated—in this example, the mother playfully moving the infant's legs—has been shown to interact with perception (Phillips-Silver & Trainor, 2005). The authors found that induced vestibular movement influences the perception and segmentation of auditory events. Taken together, this evidence suggests that the overlap of these modalities with eye contact is likely to influence the infant's attention.

The example below and the accompanying transcript further elaborate on the sequence pre-

sented above.

Transcript 5

VP03 1ˢᵗ visit (00:25-00:40)

```
1 M: (2.4 s) ((inbreathe)) geht's dir gut?(1.2 s)
  ((inbreathe)) how are you? (1.2 s)
2 M: ((makes sound smacking lips))
3 I: ((soft vocalization))
4 M: hm? (2 s)
  huh?(2 s)
5 M: alles gut? (0.9 s)
  all well? (0.9 s)
6 M: ist alles gu:t? (1.6 s)
  is everything we:ll?
7 M: ((kisses infant's feet))
```

After a pause in line 1, the mother asks the infant if she is doing well. She does this in a marked way, initiating her utterance by inhaling emphatically in preparation (Figure 5.3 and Figure 5.4 image b). At the same time, she supports her verbal behavior by temporally overlapping it with her body movement. More specifically, as shown in Figure 5.3, she literally frames her utterance with facial expressions. That is, the use of facial expression begins before and changes after the utterance. Moreover, she intensifies this multimodal input by adding head movement, another modality. As indicated in Figure 5.3, this movement is synchronized perfectly with the duration of her inhalation a utterance onset.

What follows are snapshots from the video (Figure 5.4). In (a), the face is already active; the eyebrows are raised, and a smile is present. In (b), the head moves to the back in synchrony with the inhalation; and in (c), the head moves to the front in the initial position and the face remains active.
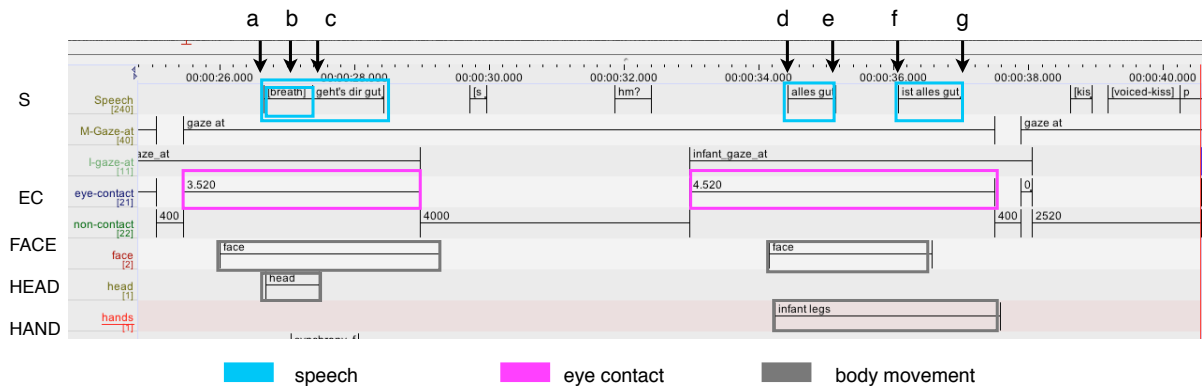
Figure 5.3.: Detail from the ELAN transcript. The mother's speech (S) is marked (intervals in light blue). Also marked is the body movement (intervals in grey) she engages in when producing these utterances within the episode of eye contact (EC) (intervals in magenta) with the infant. The letters (a-g) refer to the snapshots of the video presented below.

Figure 5.4.: Mother framing the utterance with head movement. The letters under the images indicate the exact point in time at which this movement occurs in Figure 5.3.

Figure 5.3, also shows that the facial modality continues to be active for a short time when the infant has left the eye contact episode. This indicates how the mother dynamically adapts her behavior to the infants' behavior. She realizes that the infant is no longer looking at her and adopts a more neutral facial expression.

In line 2, there is a potential attempt by the mother to attract the infant's attention, because she produces a loud smacking sound with her lips. This indicates that, in this sequence, the mother's aim is to establish or maintain eye contact with her infant; it provides evidence that the mother considers eye contact to be a state she wishes to reinforce. Her attempt is not successful this time. Next, the infant returns her gaze to the mother and eye contact is reestablished. The mother again activates her face, thus responding to the infant's gaze by modifying her behavior compared to that before eye contact. She speaks again (lines 5 and 6), repeating the same question asked before. At the same time, and while still holding her facial expression, she accompanies her utterance by touching and moving her infant's legs. Her first utterance (line 5) occurs in tight temporal relation with a first action on the legs (Figure 5.5).

(d)   (e)
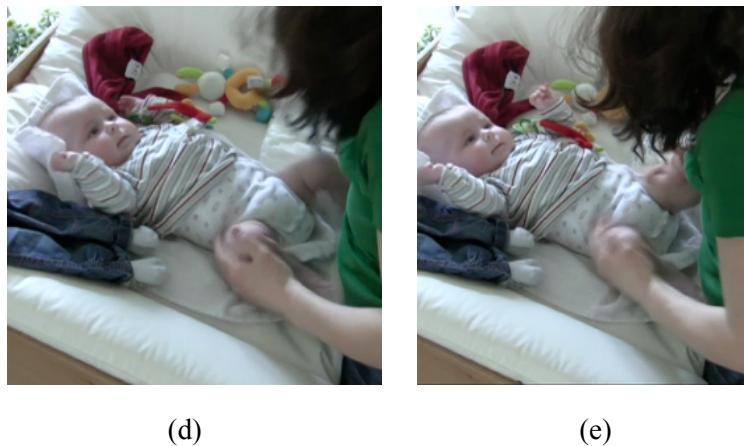
Figure 5.5.: The mother playfully caresses her infant's legs in synchrony with her speech. In Image (d) she positions one hand on the infant's leg; in (e), she moves her hands up and down the infant's thighs in an alternating movement. The letters refer to Figure 5.3.

She places one hand (d) before beginning her utterance, and then, while asking the infant, *"everything alright?"* she moves both hands up and down the legs in an alternating movement. A pause follows, and then she repeats the same question (line 6) while initiating a new action with the hands on the infant's leg. The mother now accompanies her verbal behavior by lifting up the infant's legs (see Figure 5.6).

(f)   (g)
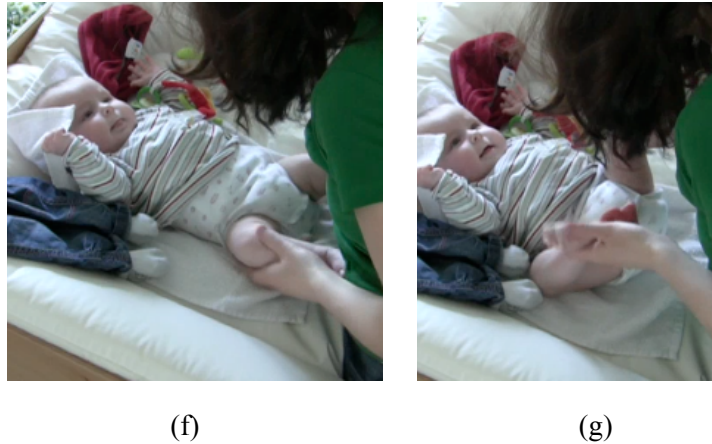
Figure 5.6.: The mother grasps her infant's legs and lifts them up. The letters under the images refer to Figure 5.3.

This tight temporal coordination of auditory stimulation with visual and/or tactile sensory information is coupled to the infant's gaze behavior where it frames episodes of eye contact. Because it has been suggested that the perception of synchrony is a mechanism guiding infants' selective attention (Bahrick et al., 2004), this multimodal synchrony co-occurring with the infant's gaze at the mother could be a powerful cue educating infants' attention: It may trigger the infant's interest in engaging in eye contact, because eye contact gets the infant involved in participating in interesting and stimulating multimodal experiences. However, it is not clear whether eye contact is established because of multimodality or vice versa. The suggestion here is that these maternal practices systematically co-occur, and, when taken together, may act as a reinforcer that frames episodes of eye contact and sets them apart from episodes without eye contact, thereby stabilizing them as a desired interactive pattern (Cowley, 2011).

Furthermore, the example above reveals not only repetition, but also variability in the mother's verbal behavior in the *"are you well?"* theme, because the mother changes the word order and uses different modalities to accompany each variation. It has been suggested that this dynamic dosing of stable and variable behaviors regulates the interest and arousal of infants (Stern, Beebe, Jaffe, & Bennett, 1977). The suggestion here is that because it overlaps with eye contact, this practice may further reinforce the infant's preference for and active engagement in eye contact. Repetitions elicit the expectation that something will follow and thus prolong engagement; variation through multiple modalities violates this expectation, thus keeping infants from habituating and looking away.

The way in which multiple modalities are organized within episodes of eye contact is also illustrated in the following example.

Example 7

VP12 ˢᵗ visit (01:36-01:44)

Transcript 6

```
1 M: (1.7 s) so: und dann ist der Part schon wieder=
   =erledigt (1.1 s)
    (1.7 s) so: and then this part is again already done
2 M: und? (0.4 s) alles gut? (0.9 s) ja?
   and? (0.4 s) all well? (0.9 s) yes?
```

Figure 5.7 illustrates the movement of the body (torso and hands) occurring within episodes of eye contact.

Once eye contact is established, the mother reduces the distance between her and the infant by moving her torso forward and looming over the infant. This is depicted in Figure 5.8. Zukow-Goldring (1996) has pointed out that looming action (moving an object toward the infant) is common to facilitate the object's perception. Gogate et al. (2000) have also found that mothers used similar forward movements of objects while synchronously naming them. Thus, the
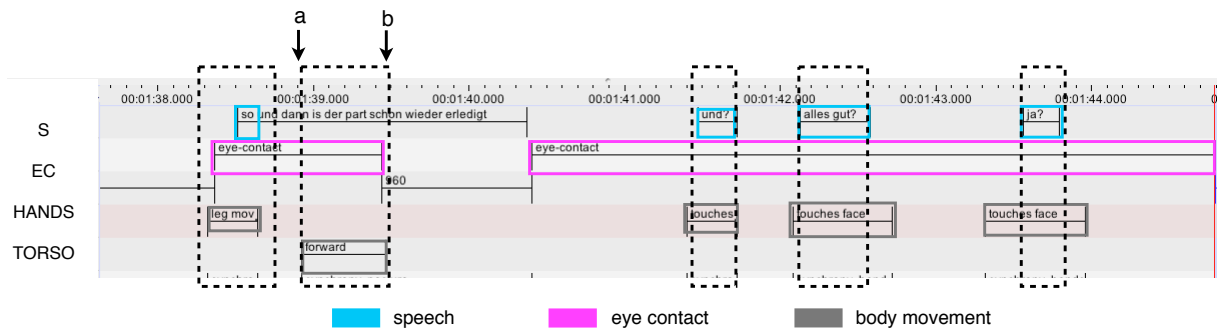
Figure 5.7.: Detail from the ELAN transcript. The mother's speech is marked (S; intervals in light blue). Also marked is the body movement (intervals in grey) she engaged in when producing these utterances within the episode of eye contact (EC; intervals in magenta) with the infant. Marked in the black boxes are also the chunks of overlapping multimodal behavior separated by pauses. The letters (a and b) refer to the snapshots of the video presented below.

forward motion seems to be especially appropriate for attracting the infant's attention.

(a)   (b)
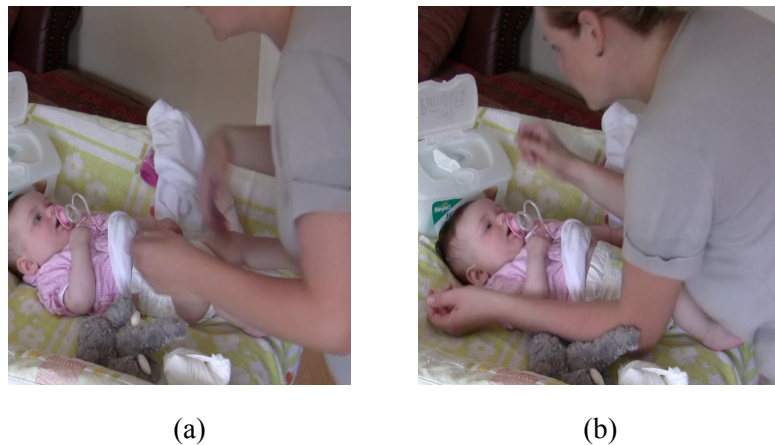
Figure 5.8.: Mother looming over the infant with her torso. The letters refer to Figure 5.7.

Furthermore, in this example, there is a tight temporal relationship between vocal behavior and body movement. When in an episode of eye contact, the mother synchronizes her questions

(line 2) with three synchronous caressing movements on the infant's cheek separated by clear pauses both in speech and movement (also marked in the boxes in Figure 5.7). This chunking of the input into short intervals separated by pauses has been suggested as assisting infants in processing the input by providing it in short bouts (Stern et al., 1977).

Summarizing the examples presented so far, the data provide evidence that mothers mark the episodes of eye contact, verbally, with greetings or questions, as well as multimodally, with movement in the visible (to the infant) parts of the body and tactile stimulation of the infant. Furthermore, the maternal practices within episodes of eye contact were repetitive. This is a general observation made across most dyads. Often, it seems that mothers—once they establish eye contact—just repeat again and again whatever it was that they were doing when the eye contact was established. Thus, in the examples mentioned, the mothers repeat questions separated by pauses and also repeat movements, for example, touching the infant's legs or cheek. It is speculated that this might be a strategy by which mothers attempt to maintain their infant's attention by producing a predictable behavior. In effect, the infant might be more inclined to look at this behavior, because she or he can validate the predictions of this interactional event (Clark, 2013). In this sense, a repetition might provide a stronger representation of an interaction pattern and be rewarding for infants' memory processes. Furthermore, the observed repetition may also involve a gradual escalation from one repetition to the next as shown in the example below.

Example 8

VP10 4th visit (14:10-14:30)

Transcript 7

In this sequence, the mother has been trying to dress up the infant, but he has been resisting. In line 1, the mother succeeds and says *"thank you"* looking back at the infant and realizing that the infant is looking at her.

```
1 M: ((sigh)) (0.3 s) dankesch n
     ((sigh)) (0.3 s) thank you
```

(a)   (b)   (c)

Figure 5.9.: Snapshots referring to transcript line 2.

```
2 M: *a(0.6 s) *b((inbreathe)) dan-ke-*cschö:n
   (0.6 s) ((inbreathe)) than-ke-you:
```


(d)

Figure 5.10.: Snapshot referring to transcript line 3.

```
3 M: *d(1.1 s)
```

(e)　　　　　　　　　　　(f)　　　　　　　　　　　(g)

Figure 5.11.: Snapshots referring to transcript line 4

4 M: *e((inbreathe)) da:n-*fke-*gschö:n

   (1.1 s) ((inbreathe)) tha:n-ke:-you:

What follows in lines 2-4 is a repetition of the mother's utterance which gradually escalates both in vocal aspects—she prolongs the syllables and over-articulates her speech—and in nonverbal aspects as she uses facial expressions, head movement (image b), and forward torso movements (image c). During the pauses, the mother returns to her starting position (images a and d). In the second repetition, she adds yet another modality to enrich her behavior, namely tactile stimulation, as she synchronously lifts the infant's legs (image e) and moves them to the both sides, bending over the infant again (image f) even further than in the first repetition (image c). Through this sequence, she succeeds in maintaining the attention of the infant. This escalation or gradual addition of new modalities may function by building up suspense and creating a playful anxiety, maintaining the infant's visual attention.

5.3.2. Directing Attention

The examples provided up to this point describe maternal practices in maintaining attention. However, what has not been considered are time periods in which the infant is not looking at

the mother—does the mother recruit the infant's attention, and if so, by what means? In the data, the general observation was that in order to recruit attention, the mothers looked for an opportunity to couple their behavior to that of the infant. Rather than attempting to change the focus of their infant's attention, more often they tried to enter their infant's attention and guide it toward them. Instead of leading their infant's attention, they chose to follow it (Tomasello & Farrar, 1986). It has been suggested that this contingent, responsive behavior encourages vocal imitation and vocal learning (M. Papoušek & Papoušek, 1990), and is thus pivotal for language acquisition, because maternal responsiveness has been found to predict the timing when linguistic developmental milestones emerge (Tamis-LeMonda et al., 2001). The following examples will elaborate on this responsiveness.

Example 9

    VP12 1st visit; (01:37-02:02)

Transcript 8

(a)   (b)
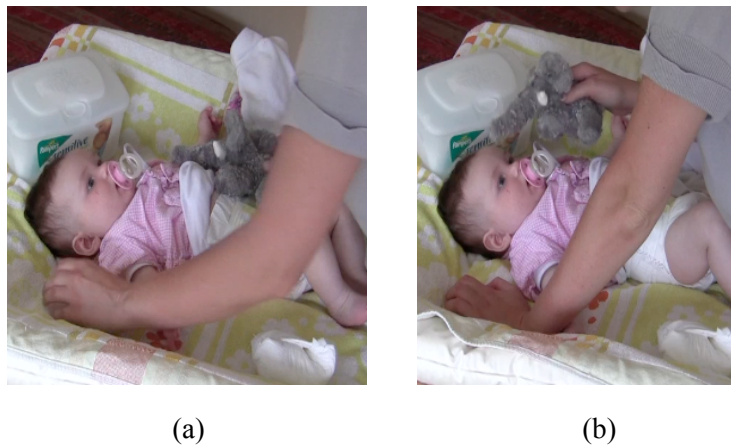
Figure 5.12.: Snapshots referring to transcript line 1

```
1 M: kennst ja nicht die *ªUmgebung (1.2 s)¹ *ᵇhm?
   you don't know the surroundings (1.2 s) huh?
```

---

[1]The underlined segments represent the intervals of eye contact.

(c)   (d)
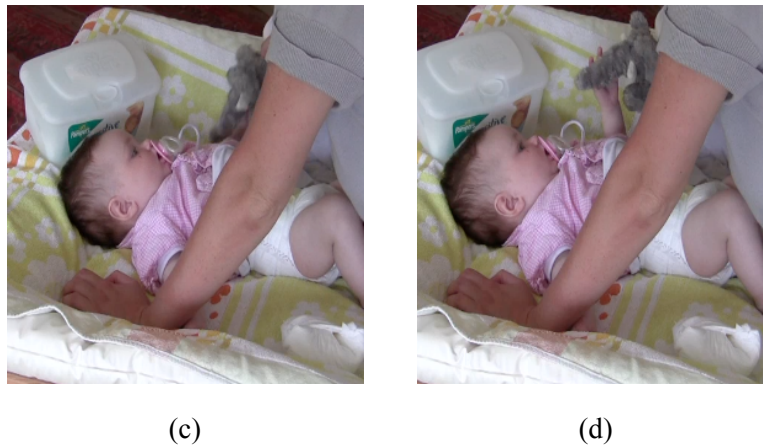
Figure 5.13.: Snapshots referring to transcript line 2

```
2 *c, *d(4.1 s)

3 M: willst du jetzt gar nichts erzählen? (3.2 s)
```
   don't you want to say anything at all now? (3.2 s)

In this example, the mother is holding a stuffed animal toy in her hand (image a), and during initial eye contact, she touches the infant on the face (image b). She then moves the toy to the side and continues to address the infant, but the infant turns to the side where the toy was placed and is no longer attending to her mother (image c). To refocus her infant's attention, the mother moves the toy toward the infant, allowing her to feel it (image d), and gradually moves the toy again in front of her (images e and f). The infant, thus, returns her gaze to the mother and the episode of eye contact is established again.

(e)   (f)
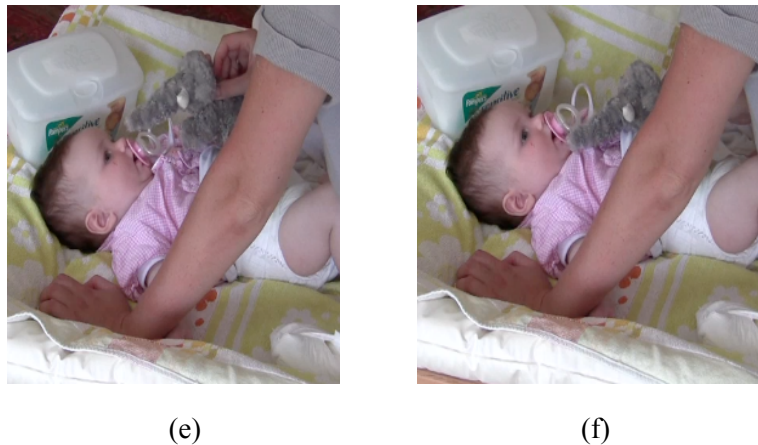
Figure 5.14.: Snapshots referring to transcript line 4

```
4 M: sonst erzählst du doch immer *eso viel (0.7 s)=
  = *f beim Wickeln
  normally you tell so much (0.7 s) during diaper changing
```

In this example, the mother tracks the interest of the infant and manipulates the object, which appears to be interesting to the infant at that particular moment. This way, she captures her infant's attention and gradually moves it back to her face. She does this by positioning the toy in her line of gaze at the infant (image e). Here, the infant can attend to both the mother and the object. After having reestablished eye contact, she puts the toy down on the infant's chest (image f). In this example, we can also see how the mother makes use of the tactile modality to direct her infant's attention: She uses a soft stuffed animal and caresses the infant's face with it. She also enables the infant to touch the toy, thus engaging her, but keeps control over the experience by manipulating the toy herself. This behavior fits well with the infant's reactivity to touch in the first months after birth; Stack and Muir (1992) have suggested that tactile stimulation acts as a social cue for infants and modulates their attention.

The following example depicts a similar strategy, but with an older infant.

Example 10

VP09 4th visit; (03:54-04:14)

Transcript 9

(a)

Figure 5.15.: Snapshots referring to transcript line 1.

```
1 M: *aräumen wir gleich weg
```
  we shall put this away in a moment

In this example, the infant can sit upright on her own. In such a position, she is naturally inclined to look down, making it more difficult to establish eye contact (image a). To redirect her infant's attention, the mother opens the sequence with a greeting (line 2), while synchronously caressing the infant's waist (image c).

```
2 M: *b(0.5 s) *challo du
```
  (0.5 s) hey you
```
3 I: ((sigh))
```

Thus, she is using a modality that seems to be close to the infant's gaze direction.

Figure 5.16.: Snapshots referring to transcript line 2.

Figure 5.17.: Snapshots referring to transcript line 4.

```
4 M: guck mal *dhier (0.2 s) Mäuschen? (.) guck mal nach oben*e
   look here (0.2 s) little mouse? (.) look up
5 I: ((vocalization))
```

She then tries again by grasping the already extended infant's arms (d). The infant does not look up. Next, she puts her hands together and positions them in an extension of the infant's gaze (e) while using a directive (line 4).
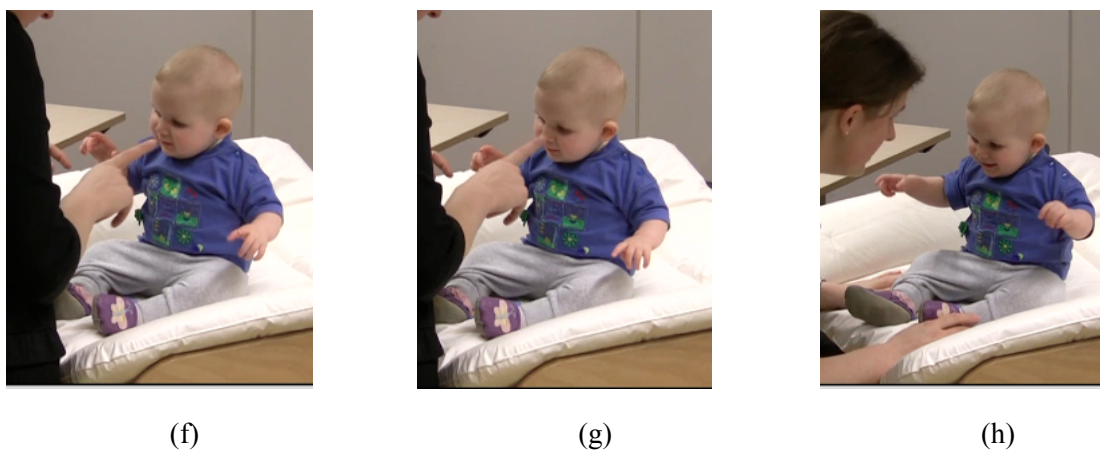
Figure 5.18.: Snapshots referring to transcript line 6.

```
6 M: hier ist die *fmami (.) *gtipsh*h
```
here is mommy (.) tipsh
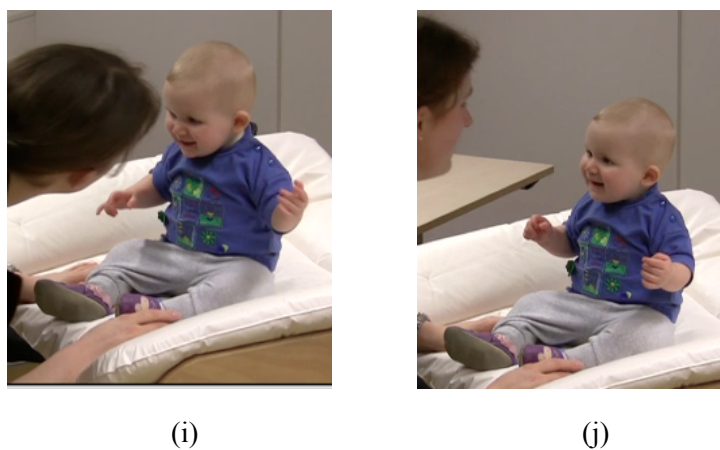
Figure 5.19.: Snapshots referring to transcript line 7.

```
7 M: *idibidibidibidibidibidush (0.2 s) <inbreathe>*j=
  =dibidibidibidibidibidupsh
8 I: ((vocalization))
```

Moving her fingers upward, the infant follows the movement of the mother's fingers and looks up. Having entered her infant's attention, the mother quickly lowers her body and positions her face at the level of the infant's gaze (image h) while at the same time saying *"here is mommy"*. When eye contact is established, the mother starts to hover over and repeatedly loom in toward the infant to maintain it (images i and j).

Here, again, the mother is using the infant's preference as a point of reference around which she weaves her own behavior in order to recruit the infant's attention and guide it to the establishment of eye contact. She uses both visual modalities—moving her fingers playfully in front of the infant's face (image f)—as well as tactile stimulation (images c and g), and additionally coordinates her vocal behavior with her body movement to provide a multimodal sensation.

Another way of coupling the behavior of the infant for eye contact is the imitation of infant vocalizations, as the following example elaborates.

Example 11

VP10 4th visit (02:50-03:00)

Transcript 10

```
1 M: ((inbreathe)) ueber die Schulter
     ((inbreathe)) over the shoulder
2 M: (0.9 s) und die andere Schulter
     (0.9 s) and the other shoulder
3 I: ((click sound))
4 M: ((imitation)) ((imitation)) ((imitation))
```

In this sequence, the mother is massaging the infant's body. She is looking at the infant and verbalizing her actions (lines 1 and 2), but the infant is looking away. In line 3, the infant produces a soft click sound. The mother reacts rapidly (0.3 s) and imitates the sound the infant just produced in a more exaggerated manner, that is, much louder and three times in succession. The infant reacts by looking back at the mother and establishing eye contact (Figure 5.20).

Figure 5.20.: Detail from the ELAN transcript. Marked is the infant vocalization (S-I), mother imitation (S-M), and infant gaze at mother (IG).

This example illustrates a different kind of adaptive behavior by the mother. This time, instead of moving into the infant's visual attention, she reciprocates his behavior. In this case, this proved to successfully recruit the infant's attention. This is consistent with research suggesting that imitation of infant vocalizations motivates infants to engage actively in early 'conversational' games of vocal matching (M. Papoušek, 1995), thus capturing their attention. After establishing eye contact, mother and infant playfully move in and out of eye contact while vocalizing in alternation.

Summarizing the observations presented above, it was shown that recruiting attention involves multimodal practices that are co-constructed in the moment-to-moment flow of interaction and are responsive to the infant's interest and focus of attention.

5.3.3. Pursuing eye contact

Example 12

VP10 4th visit (07:30-07:43)

This sequence is the last in this particular interaction. The mother has finished diapering and

marks the end by saying *"finished"*, elongating the vowels in the utterance (line 1) and making a large gesture with her hands (image b and line 2).

Transcript 11

(a)

Figure 5.21.: Snapshots referring to transcript line 1.

```
1 M: *ªfe::rtig
```

finished

(b)

Figure 5.22.: Snapshots referring to transcript line 2.

```
2 M: ((inbreathe)) fe:*ᵇrtig (1.6 s)
```

((inbreathe)) fi:nished (1.6 s)

What is interesting here is that she appears to expect a specific reaction from the infant which she does not get, and this becomes evident in her subsequent pursuit. Pursuit of a particular response has been suggested as being an educational tool used by parents in conversation with children learning how to speak (Filipi, 2013). Here it is used in the service of the education of attention.

In the entire sequence, the mother is looking at the infant's face but it seems that she cannot get him to look at her. In the beginning of the sequence when she says *'finished'*□□, the infant is looking at her hand (image a). Subsequently, she repeats her utterance, adding the hand gesture, but that still fails to attract the infant's attention to her face; he follows the movement of her hand (image b). Having no success, she makes another attempt, this time by calling the infant's name (line 3), but as marked in image (c) with the grey arrow the infant is now looking away.

(c)

Figure 5.23.: Snapshots referring to transcript line 3.

```
3 M: *c((infant name))? (0.6 s)
```

In her next pursuit, she taps the infant on the belly, while calling him by his full name this time (line 4 and image d). She first succeeds in establishing eye contact at this moment (image d), and so she returns to her original utterance (line 5), but immediately loses the infant's gaze again (image e).

(d)

Figure 5.24.: Snapshots referring to transcript line 4.

```
4 M: ((infant name))*d (.)
```
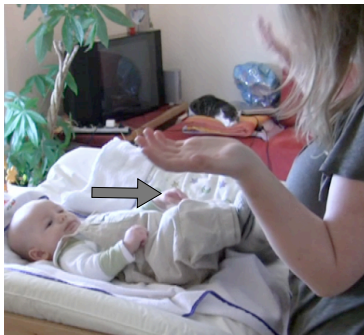

(e)   (f)

Figure 5.25.: Snapshots referring to transcript line 5.

```
5 M: *efe:rtig (0.9s)*f
   fi:nished
```

Having lost the infant's attention, she moves her upper body in his visual field (image f) and then continues with a greeting (line 6). Having recruited his attention, she repeats her original phrase, concluding with a smile (image g). This conclusion is further marked by a modified

intonation of the repeated phrase. The infant smiles too (image g), and only then does she end the interactional sequence, reaching to lift the infant from the changing table (image h).

(g)

Figure 5.26.: Snapshots referring to transcript line 6.

```
6 M: hallo (.) fe:rtig*g (- - -)*h
   hello (.) fi:nished (- - -)
```


(h)

Figure 5.27.: Snapshots referring to transcript line 6.

This sequence illustrates the way in which eye contact is dynamically negotiated. It exemplifies the process of co-shaping mutual attention. In pursuing eye contact in this example, the

mother repeatedly takes opportunities to reinforce her gaze at the infant by coordinating her vocal behavior with her body movement. This is also apparent in Figure 5.28. The majority of the mother's verbal behavior overlaps temporally with movement of her body, i.e., hand movements (images b and d), facial expressions (images f and g), tactile stimulation of the infant (image d), movement of the head (image f and g), and movement of the torso (image f and g).
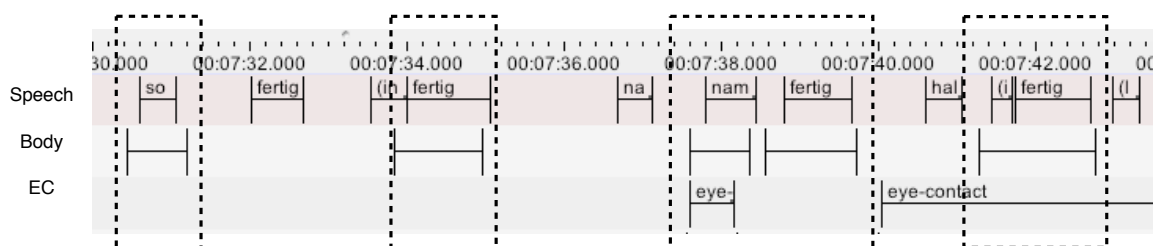
Figure 5.28.: Detail from the ELAN transcript. The tiers represent, from the top, the mother's verbal behavior, her body movement, and the eye contact between mother and infant. The chunks of multimodal behavior are marked in the boxes throughout the sequence.

This sequence further exemplifies how the education of attention might be tuned both during the course of a particular interaction sequence as well as over the interactional history of mother and infant (Fogel et al., 2006). In this sequence, it becomes apparent that this dyad has established a repertoire of interactional formats, so that not just any response of the infant is arbitrarily treated as relevant, but that a specific action sequence is expected for the interactional episode to be successful. In this case, mother and infant have to mutually acknowledge the termination of the interaction by looking and smiling at each other.

5.3.4. Comparison of Age Groups

So far, multimodal practices of maintaining, recruiting and pursuing eye contact have been discussed and have been identified in examples from both age groups. In the following, the two age groups shall be compared.

In the chapter on gaze patterns, the topic of 'gaze-role reversal' was already addressed. It was indeed a striking observation that a comparison of three- and six-month-old infants seemed to reveal a role reversal concerning who is the one (mother or infant) maintaining a more constant gaze and who is the one moving in and out of eye contact. As already illustrated in chapter 4, the observation was that for the three-month-olds, it was the infant who was looking at the mother's face for long periods, whereas it was the mother who set the pace for the organization of eye contact. It was she who exited the mutual gaze episodes to continue with her activity of diapering the infant, that is, wetting her cleaning cloth and cleaning the infant's skin. When she looked back at the infant, the infant was already looking at her. In observations of the mother-□infant dyads when the infants were six months old, the gaze pattern had changed. The infants no longer framed the gaze episodes. They engaged more actively in quickly switching their gaze to and away from the mother.

Also in this previous chapter, the quantitative analyses supported this qualitative observation. The six-month-old infants spent significantly less time looking at their mother's face and significantly more time looking at her hands, body, objects, and the surroundings. Also, the duration of infants' gaze events at the mother became significantly shorter. So there is a development in the gaze patterns. Also it was shown that the mothers' average durations of gaze events decreased significantly for gazing at the infant. In the previous chapter, this was discussed as representing the adaptive maternal behavior found in the microdynamics of gaze organization: As the infant becomes more skilled and faster at controlling his or her gaze, the mothers adapt their gaze behavior to that of the infant.

The question arising from this observation is whether this development of infant attention influences the manner in which the maternal multimodal practices are employed and if so, how?

The first adaptation observed was of a temporal nature. As the six-month-old infants infants have significantly reduced the duration of gaze events at the mother, the mother now needs to actively search for slots of eye contact to provide responsive reinforcing eye contact. Here the observation was that the multimodal practices become very concentrated and short as illustrated in the following example.

Example 13

This example is from a sequence already analyzed. Below an excerpt from Transcript 3.

VP17 4th visit (00:16-00:19)

(a)   (b)   (c)
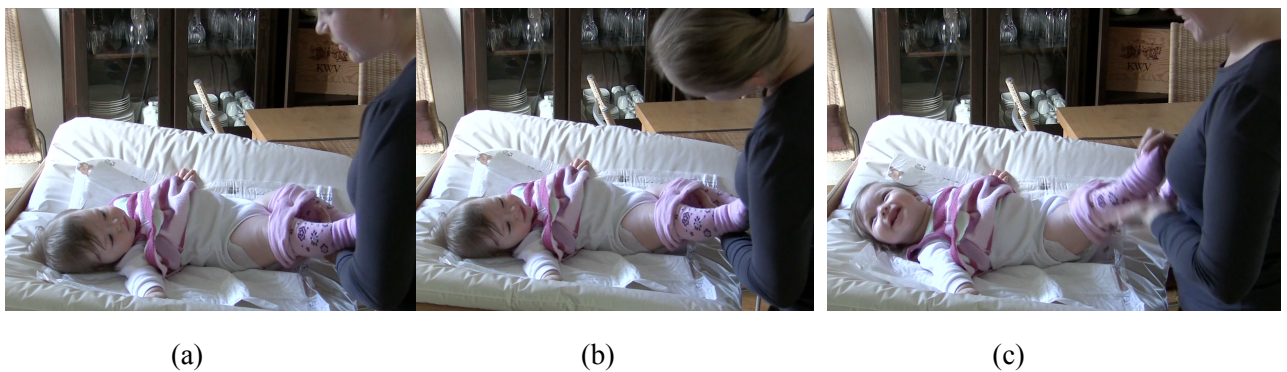
Figure 5.29.: Snapshots referring to Transcript line 4

```
4 M: *akuck kuck *b, *c(4.5 s)
```

In this example, the infant is initially looking away (image a). She turns her head to face the mother, establishing eye contact (image b). The mother quickly moves her upper body sideways to match the position of the infant's face orientation (image b). She simultaneously provides a verbal utterance saying, *"Kuck Kuck"*. The infant smiles and turns away again (image c). In this example, the mother has very little time to provides her multimodal behavior, as the infant's gaze at the mother's face does not remain for long. This is different from the observations of the three-month-olds, in which the mother has more time to organize her behavior.

Example 14

This example is based on a sequence already presented in a previous section.

VP03 1st visit (00:25-00:40)
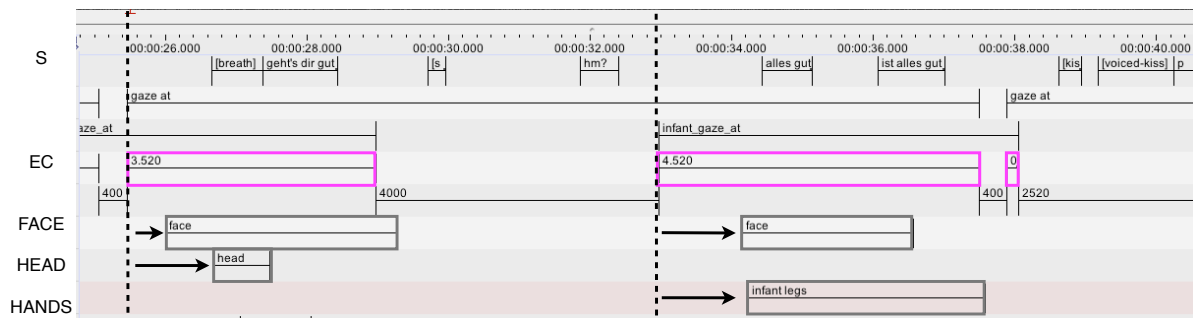
Figure 5.30.: Detail from the ELAN transcript. The mother's body movements (intervals in grey) are coded for face, head and hand movements. Episodes of eye contact (EC) (intervals in magenta) with the infant are also marked. The vertical lines mark the onset of eye contact and the arrows the delay with which the mother initiates the multimodal practices.

In this sequence the eye contact window is to a great extent longer than the duration of the multimodal practices (See Figure 5.30). Here, the face, head, and hands modalities are activated within the eye contact window and not directly at the onset of eye contact. As shown in Figure 5.30, there is a delay in the onset of these modalities. This gives the impression of a more gradual and slower use of synchrony within eye contact in the interactions with the three-month-old infants.

In the comparison of age groups another form of adaptation involves the modalities used in establishing and maintaining mutual attention. There was a striking decrease in infants' attention to the face of the mother and an increase in attention to her hands and objects, which was also confirmed by the quantitative analysis of gaze patterns reported in the chapter on gaze patterns. Compared to the three-month-olds, mother and infant spent far more time during the second

visit touching each other and touching objects while engaging in eye contact. Thus, there seems to be a shift in the modalities used for recruiting and maintaining attention in the two age groups as elaborated in the following examples.

Example 15

VP14 1ˢᵗ visit (04:24-04:30)

In this sequence, the mother is looking at the infant while the infant is looking to his side. After about one second, mother and infant establish eye contact (see Figure 5.31).
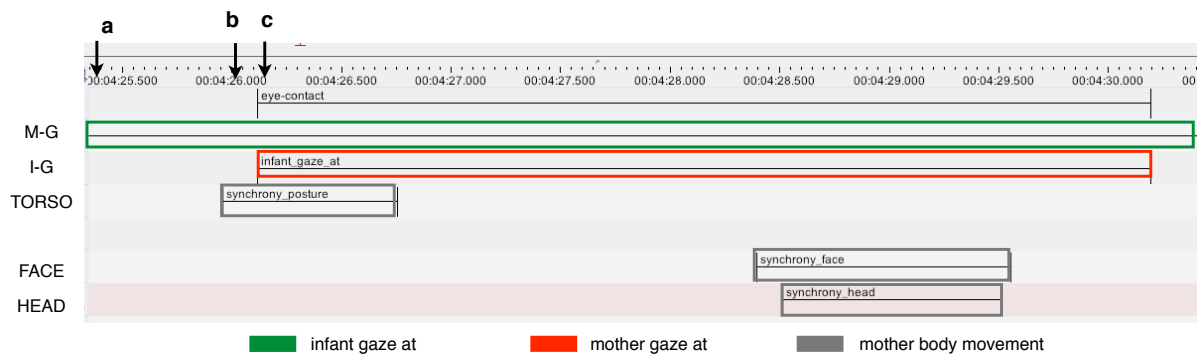
Figure 5.31.: Detail from the ELAN transcript. The mother's body movements (intervals in grey) are marked for torso, face and head movements. The infant's gaze at the mother (I-G) (intervals in red) and the mother's gaze at the infant M-G (intervals in green) are also marked. The letters (a-c) correspond to the snapshots presented in Figure 5.32

In image (a), the mother is touching the infants' feet but the infant is not responding with gaze. To establish mutual attention, the mother bends forward (image b) while at the same time calling the infant by his name. The infant turns (image c) and looks at the mother. Once eye contact is established, the mother maintains it with further multimodal behavior: After a pause, she produces an inbreathe and a greeting, synchronized with movements in her face and head movement. In this sequence, the mother is using modalities which focus on the area of the face: looming movement, eyebrow movement, smile and head nod.

(a)   (b)   (c)
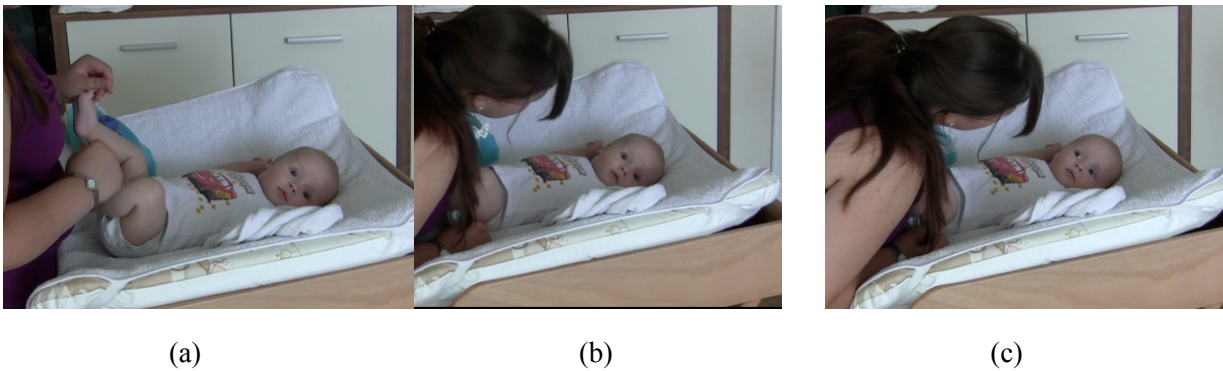
Figure 5.32.: Forward movement of the mother to recruit attention. The letters under the images correspond to the positions marked in Figure 5.31.

Below, an example of the same mother-infant dyad three months later is presented.

Example 16

VP14 4th visit (01:03-01:07)

(a)   (b)
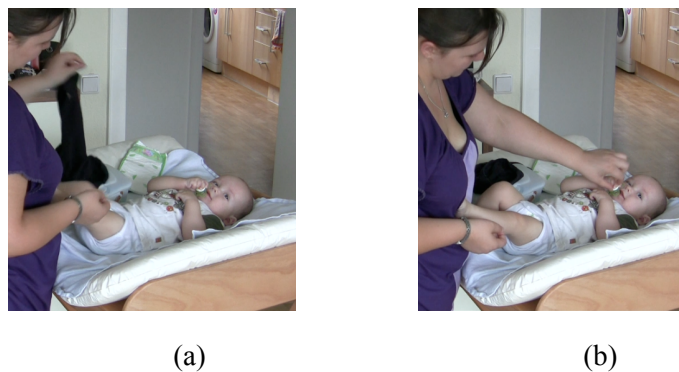
Figure 5.33.: Recruiting attention using the tactile modality. Part 1

In the beginning of this sequence, illustrated in Figure 5.33, the infant is not looking at the mother (image a). He is holding his pacifier, and is looking to the side. To establish mutual attention, the mother reaches and touches the infant's hand and his pacifier (image b). She slightly lifts them; at this moment the infant turns and looks at her face (image c). The mother playfully wiggles the infants' hand and pauses, releasing it (image d).

(c)   (d)
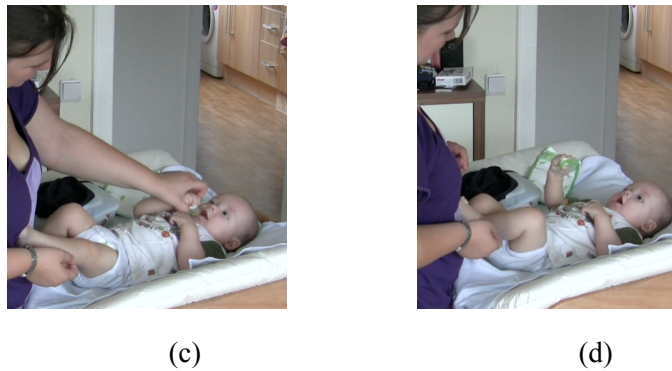
Figure 5.34.: Recruiting attention using the tactile modality. Part 2

As eye contact has now been established, the mother continues to rhythmically grasp and release the infant's hand and the pacifier, while as the same time repeating *'it's mine, it's mine'*. In this example, the mother does not choose a visual modality but a tactile one. Whereas in the previous example, the hands modality was unsuccessful in recruiting and maintaining the infant's attention, after adding the movement of her torso, face, and head, she succeeded. This example shows the changes in the development of the education of attention. In the earlier months of the infant's life, to elicit responsiveness the mother is required to use her torso and her face, as her hand movement is insufficient. In the fourth visit, we can observe a qualitative shift in the modalities used; the infant's attention now can be recruited with the mother's hands.

5.4. Summary of findings

The present data analysis addressed the question whether and how the development of mutual attention is supported, and thus educated, in early interactions through social reinforcement. For this purpose, the analysis aimed at exploring whether eye contact, as one early instantiation of mutual attention, is reinforced behaviorally in natural mother-infant interactions, and how this (1) goes hand in hand with the development of the infant and (2) changes as a function of the mother and infant's joint interaction history (Flom, Deák, Phill, & Pick, 2004).

The focus was on the sequential organization of gaze behavior of the mother and the infant, with the aim to elucidate maternal strategies in recruiting or maintaining eye contact with her infant. The analysis showed that eye contact was reinforced systematically by accompanying verbal and nonverbal behaviors.

More specifically, the data revealed that eye contact episodes overlapped with verbal behaviors that included opening sequences such as greetings, questions, or calling the infant's name—behaviors that have been suggested to have ostensive power, because they signify to the infants that something is explicitly designed for them and addressed to them.

It was shown that eye contact systematically occurs simultaneously with other social signals. Furthermore, the analysis showed that the verbal behavior accompanying eye contact overlapped not only with the mother's body movements such as facial expressions, head movements, torso movements, and hand movements, but also with tactile stimulation of the infant. These multimodal behaviors were performed in a very tight temporal relationship to one another, thus forming multimodal chunks of behavior that frame episodes of eye contact. In these chunks, it was shown that behavior occurred systematically when the eye contact was established and less frequently when eye contact was lost.

To maintain attention, mothers used repetition of the multimodal packages of stimulation, which they clearly separated from each other with pauses.

In addition, the analysis of interaction sequences revealed that in order to recruit attention, the multimodal input provided to the infants during eye contact was triggered interactively. Thus, the infant has an active role in co-constructing the ostensive input that she or he receives, because the mother uses the focus of her infant's attention as a *'hook'* for her contingent behavior.

Concerning the development of mutual attention and multimodal practices over time, the longitudinal analysis revealed differences in both the organization of gaze as well as multimodal practices: In accordance with the findings presented in the previous chapter on the gaze patterns of mother and infant, it was shown that the multimodal behaviors of the mother were adapted to the faster gaze switches used by the infant and also grew shorter and increased in intensity.

Additionally, a change in the modalities used was observed.

These observations raised new questions concerning the exact kinds of multimodal practices used by the mothers in the data corpus and about their use in the interactions with their three and later the six-month-old infants. The investigation of these questions will be described in the next chapter.

6. Synchrony between language and action

6.1. Objective

In the previous chapter it was shown that mothers in their interactions with both their three-month-old and six-month-old infants often use multimodal practices. In the explorative analysis, it became apparent that mothers coordinate their verbal behavior with their body movements, and that this coordination occurs in a very tight temporal overlap. The aim of this chapter is to look into this temporal coordination with micronanalytical precision. To find out whether and how language and action are coordinated in early mother-infant interactions, the following questions guided the analysis:

1. What are the pragmatics of the temporal co-occurrence of language and action: During which practices do they become synchronized, and in which ways?

2. To what extent do mothers offer multimodally overlapping stimulation?

3. From which different modalities does this stimulation originate?

4. Does this multimodal stimulation change over time?

6.2. Method

6.2.1. Participants and Data

For this analysis, the data originated from the same corpus described in the analysis of gaze in the previous chapter, i.e., it consists of the videos from the first and fourth visits to the families' homes, when the infants were three and six months old, respectively. Again, all seventeen mothers and their infants were considered for this analysis.

6.2.2. Data analysis and coding

Data analysis was again twofold. First, qualitative microanalysis was used to reveal the different qualities of the intermodal co-occurrence in maternal stimulation. Then, a coding scheme was created and the data coded according to these categories. These, in turn, served to describe the data quantitatively in a next step of analysis.

The goal of the coding process was to describe maternal behavior in terms of the temporal coordination between the mothers' vocal behavior and body movement in the interactions with their infants while making use of intermodality, i.e., the interplay of stimulation originating from different modalities. The description was initially carried out by viewing the behavior of the mother as originating from her own agency, rather than from a dyadic perspective. This decision influenced the entire system of coding classes, because they were taken as intended by the mother and not as interpreted or perceived by the infant. For example, when the mother touched the child, this was coded as movement of the mother's hands and not as the infant being touched. Data were scored with INTERACT®9(http://www.mangold-international.com).

Also, the audio signal was separately coded using PRAAT phonetics transcription software (Boersma & Weenik, 2010) to measure the total duration of maternal verbal behavior.

Coding of body movement classes

Coding of the data was undertaken in a number of steps. First, all clips were viewed innumerous times at normal speed. This step served as an initial familiarization with the data. Notes were taken down, describing action patterns, repeated routines, and activities during each video, as well as observations regarding processes occurring across subjects. These notes served as the first indices and observations, which would lead to a gradual elaboration of the coding definitions. Clips were viewed one after the other for each mother-infant dyad. This aided the observation of within-dyad actions and routines. Clips were also watched alternating between individual mother-infant dyads, which helped generalize observations.

In the data, the mothers were engaged in a manual task involving constant body movement while simultaneously interacting with their infant. Thus, the data consist of two parallel running signals: a vocal signal and a motor signal.

The units of analysis were inspired by the method developed by Stern et al. (1977) for a unimodal signal and adapted for the purposes of the present analysis. The authors defined "vocal and kinesic phrases" as "discrete and shorter bursts of behavior with more obvious 'rests' in between bursts" (Stern et al., 1977, p. 181). Phrases were intervals of either vocalization or body movement that were clearly demarcated by prior or subsequent longer pauses. During these intervals, the context in which maternal vocal and motor behaviors were coordinated communicatively can be observed. Yet, in the present approach, the concept of phrases (Stern et al., 1977) was extended. In contrast to coding vocal behavior and body movement separately (Stern et al., 1977), the initially identified intervals were of a multimodal nature.

In a second step which finally led to the quantitative analysis, the exact minimal segments during which body movement co-occurred perfectly with speech were identified as temporally synchronous. In the case of ongoing action, the focus was on those moments in time when the mother marked her action with her verbal behavior. In the case of ongoing verbal behavior, the focus was on those movements of her body that were coordinated temporally with her vocalizations. These instances produced the basic units of analysis.

The following parts of the body became visible during interactions: the hands, the torso, the head, the face, and the lips. Movement in these parts of the body co-occurring with vocal behavior became the classes of body movement that are also referred to as 'modalities' in the following. A modality was annotated when:

a) Vocal behavior co-occurred with body movement

b) During the coded intervals, both verbal and motor behaviors were continuously active in the form of an ongoing vocal behavior or a continuous movement.

In the case of facial expressions, the general observation was that facial movements were subtle and fleeting. Once moving to a facial expression, some mothers 'froze' this expression, maintaining the tension in their face. They then released the expression and reassumed a neutral face. This 'freezing of the face', established a continuous state of some duration and was considered an interval of ongoing activity in the coding.

The data were coded continuously for synchrony. Videos were viewed from the beginning to the end, and the onset and offset of each modality interval was annotated. Minimal segments were those during which body movement co-occurred perfectly with speech. This means that when, for example, the onset of head movement started before the vocal utterance, the moment in which the vocalization started was marked as the onset of a segment and the moment in which one of the two modalities ceased to be active as the offset. Thus, the resulting coded intervals consist of minimal segments of synchronized vocal behavior and body actions. It might be the case that this strict coding yields small segments of behavior that seem somewhat incomplete and meaningless. For this reason, the two step coding process is so valuable: Taking the conservative coding approach, the quantitative description provides findings of perfect synchrony between vocal behavior and body movement employed in maternal input to very young infants, whereas the qualitative analysis aims to describe the meaning of the synchrony.

Separate coding runs were made for each modality since it was more efficient to concentrate on one modality at a time while going through the data. After the entire data set was coded,

annotations were viewed repeatedly in slow motion and refined for exact frame accuracy. Table 6.1 lists the coding classes. In the following, each coding class shall be described in detail.

Coding class: Hands

Both hands were coordinated with each other most of the time and were mostly engaged in movement. Therefore, instead of having two separate classes for the left and right hand, there is a single joint category for both. Coding of the hands' body class was carried out by playing the video clip at one-quarter speed or even slower and marking the parts during which movement of the (upper and lower) arms, hands, or fingers co-occurred with vocal behavior. The coded intervals were then repeatedly played back and forth, frame-to-frame, in order to fine-tune the specific onset and offset of the codings with single-frame precision.

Coding class: Upper body

The body was coded in terms of its forward, backward, and sideways movement when coinciding with vocalizations. When first exploring the data, the observation was that upper body movement occurred quite often during each session. Mothers loomed over the infant to place themselves in the focus of the children's gaze, or moved their upper body while changing the baby. For this data, only cases in which the movement of the upper body clearly coincided with vocal behavior were coded as co-occurring.

Coding class: Head

The movement of the head was observed to occur in synchrony with maternal vocalization: The data revealed exaggerated and accentuated nodding (up-down as in saying *"yes"* and left-right as in saying *"no"*) as well as sideways movements or even circular movements that seemingly followed the intonation pattern of the utterance. Coding of the head movement was carried out via frame-to-frame analysis. The onset of coding was annotated upon the change of the orientation or trajectory of head movement. Offset of the coding was annotated when the head

Table 6.1.: Coding Classes and their Description

Body Class	Description
Hands	Mother gesticulates, handles the baby, pats, tickles, or rocks the baby playfully in synchrony with her vocalizations. Mother may also manipulate objects (such as the wet wipes, the baby's clothes, or even toys) in synchrony with her vocalizations.
Torso	Mother moves her upper body in synchrony with her vocalization. Includes bending over the child, playfully moving from side to side, or even shaking her upper body.
Head	Mother nods (up-down, left-right), or merely moves her head abruptly visibly in one direction. Mother may also shake or move her head playfully from side to side when singing or speaking to the child.
Face	Exaggerated face: either of mock surprise, with raised eyebrows, or frown while breathing in loudly or sighing. Mother may also use facial expressions to highlight certain aspects of her speech.
Lips	Mother kisses the infant making a loud smacking sound, or voicing the kissing action. This behavior includes the employment of the tactile modality when the mother kisses the baby on the face or the body. It can also be produced without contacting the infant body

stopped at a specific position for longer than three frames. Coding of the head was impossible to achieve with sound, because the frame intervals were too short to understand what was being said. Thus, at a later step, all annotations were reviewed and only those accompanied by synchronous vocalization entered the analysis.

Coding class: Face

In the data, facial movements were used in a systematic way: Mothers employed exaggerated facial expressions as an extra modality to be synchronized with their vocal behavior. Furthermore, there was a tight relationship between head and facial movements. Actually, the onset of head movement could often be predicted: About one or two frames before the actual head movement, the eyebrows started moving upward. This co-occurrence and correlation has already been reported elsewhere (Kaye & Fogel, 1980).

Coding at a finer level of detail was carried out via frame-to-frame analysis. Coded changes in the mothers' faces seemed to move systematically in two opposite directions: outward and inward. Mothers were observed to make three primary facial expressions: a big, wide face, with raised eyebrows or even an open mouth of mock surprise, a big smile or open mouth during phases of positive affect (outward movement); and a small wrinkled face, with eyebrows pushed together toward the middle of the forehead in an exaggerated frown, mostly during times when the infant was showing negative communicative signs such as being about to cry (inward movement). As in the case of head movements, all data were coded for facial expressions without sound, after which, in a second step, annotations with no indication of vocal-facial co-occurrence were dropped from the analysis.

Coding class: Lips

The last modality entered into the analysis due to the observation that kissing occurred in many of the recorded sessions. Although it was produced by a part of the face – the lips – it differed from the other facial features coded in the face modality in two ways: First, the mouth has

a different expressive nature, because the lips can provide additional tactile stimulation when contacting the infant's skin. Second, during kissing, lips can also produce a sound, which makes them inherently multimodal. Kissing intervals were thus coded separately.

6.3. Results from qualitative microanalysis

6.3.1. Idiosyncratic interaction structures

During each session, and given the experimenter's instructions to act as though they were not in a hurry, mothers prolonged the process of changing the diaper as long as they felt their infant was comfortable with it (some interactions lasted as long as 17 minutes). Sessions could be divided into smaller episodes. Mothers spent some of the time on the actual task of changing the infant diaper: undressing the child, wiping the infant clean, dressing the child again, and so forth. The rest of the time, mothers spoke to the infants, played with them, sang to them, engaged in reciprocal dialogue-like interactions, and attempted to sustain an atmosphere of positive affect. Yet, each mother–infant dyad had its own routine and individual process rendering any generalization of distinct interaction phases across subjects impossible. This high interactional variability has already been stressed by Fogel (1993b) who believes that mother–infant dyads develop unique "consensual frames" of interaction, elaborated through intra-dyad "cumulative" (Hsu & Fogel, 2001, p. 1061) interaction history. For example, one mother had a clear structure in her interaction. First, she undressed the infant, then she played with him. As soon as she noticed the infant beginning to show signs of discomfort, she put a clean diaper on him and dressed him. Another mother showed a more variable interplay of behavioral routines, during which playing and changing the diaper were integrated into a whole, making a clear general structuring and subdivision impossible.

Also, there were differences between subjects concerning the maternal interactive style: Some mothers gave an impression of leading the interaction; they opened topics of 'conversation' and told stories. Others seemed to be led by the infant's behavior and imitated infants' sounds and

faces, constantly commenting on the infant's state. Concerning verbal behavior, some mothers gave the impression of being more verbal, and providing more speech-like input, with rhymes and songs, whereas others gave a more corporeal, sensory quality to the interactions, providing exploratory infant-like sounds and more onomatopoeic expressions.

Not only with respect to their style in interacting and verbalizing, but also with respect to the amount of input the mothers provided, there was great variability in the measured variables. The mothers varied also greatly both in the overall degree to which they coordinated their vocal and motor behavior, as well as in the modalities with which they coordinated it. There was even one case in which the mother made no use of the facial modality at all. Interestingly, this particular mother exhibited the highest degree of coordination of overall body movement and speech, suggesting that she compensated for her lack of facial behavior by using other parts of the body. This finding reinforces the idea that mothers cannot be classified as 'more' or 'less' multimodal.

In spite of the great variability with which the mothers executed the task, the differences in the materials they used, and the ways in which they structured the interaction, there were certain behaviors that could be found across subjects. By closely observing the ways in which vocal behavior was coordinated with body movement, practices were discovered that formed behavioral patterns common to most of the mothers participating in the study. These are described below, with specific examples from the data.

6.3.2. Coding class: Hands

A synchronization of language with the hand movement was observed in a variety of ways. During the actual manual task of changing the diaper, mothers used their hands to dress and undress the infant, clean the infant, and move, turn, and lift him or her. Such an activity was often notably accompanied by vocal stimulation. Mothers mostly spoke about their actions, that is, what they were in the process of doing, and their language accompanied each action step, as shown in Example 17.

Example 17

VP14 1ˢᵗ visit (03:07–03:15)

Transcript 12

(a) [introduces action]   (b) [action step one]

Figure 6.1.: Snapshots referring to transcript line 1

```
1 M: jetzt *apacken wir alles ei::n*b (1.2 s)
   now we pack everything in (1.2 s)
```


(c) [action step two]   (d) [action step three]

Figure 6.2.: Snapshots referring to transcript line 2

```
2 M: erst die ei:ne Seite*c (0.5 s) dann die a:ndere *dSeite, ne?
   first one side (0.5 s) then the other side, huh?
```

In this example, there is a tight temporal relationship between what is being said and what is being done: The mother is segmenting her behavior by vocally demarcating each intermediate step of her action sequence. She is also extending the length of certain words, thus highlighting the duration of each movement, providing for synchrony at the conclusion of each action step. Thus, in this case, language is being adapted to match with and structure the actions performed.

There were also other ways of marking parts of an action: The beginning of an action was commonly accompanied by loud breathing, or by counting *"one two three"* functioning as an ostensive cue to announce the initiation of an action. This is illustrated in the following example.

Example 18

This is the final sequence in this interaction. The mother has finished diapering and dressing the infant and is about to pick him up from the changing table. To do this, she leans forward (image a) to lift the infant. As she puts her hands on the infant's body to perform this action, she synchronizes her action with a loud inhalation (image b). She then lifts the infant, marking the duration of this action with the interjection *'yi::ippy'* (image c).

VP16 4th visit (02:41–02:43)

Transcript 13

| (a) [action preparation] | (b) [action announcement] | (c) [action completion] |

Figure 6.3.: Snapshots referring to transcript line 1

```
1 M: na *a komm *b((inbreathe)) ji::ieppie*c
   come on ((inbreathe)) yi::ippy
```

In this example, it seems that by synchronizing the onset of a new action with an inbreathe the mother is communicating to the infant that *"something is going to happen shortly".* This is supported by the observation that this practice occurred mostly when the mother was about to act on the infant's body in some way, for example turn him on his the side or lift him up. It seems as if the mother was giving the infant notice that something is about to happen to him or her.

Some other practices were common to mark the end of an action, such as the use of interjections and onomatopoeic words expressing sounds.

Example 19

In this example, the mother marks the effective end of her action of putting on the infant's trousers with the interjection *"plup"* after pulling up the trousers and letting the infant fall back playfully on to the changing table.

VP03 1st visit (04:28–04:30)

In image (a) the mother places her hands on the infant's body. In image (b) she lifts the infant's lower body, and in image (c) she releases the infant's body marking it by saying *"plup"*.

Transcript 14

(a) [action preparation]   (b) [action duration]   (c) [action completion]
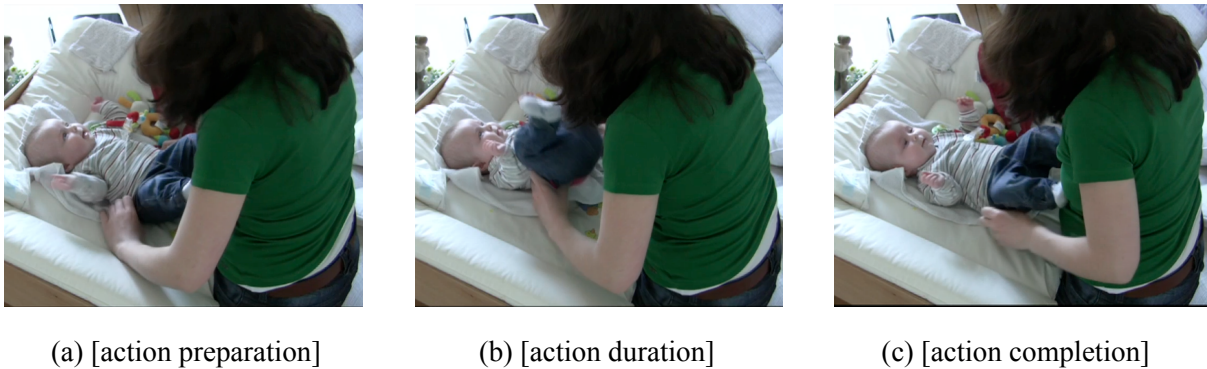
Figure 6.4.: Snapshots referring to transcript line 1

```
1 M: *a *b plup*c
```

Similar examples are ample in the data. For example, another mother used the interjection *"tsak"* when closing the package of wet wipes. Also, single words were used to segment an action. For example, one mother accompanied different movements with *"one, two, three"*; others spoke at the end of a movement. Thus, in these cases, there was only a partial overlap of the action with the vocal activity that highlights the goal of the action.

A synchronization between the two modalities (hands and language) could also be observed in some mothers who highlighted the duration of actions, by expanding the length of their verbal behavior, thereby matching the duration of their speech with that of heir movement. For example, one mother prolonged the word *"and"* accompanying the action of grasping a sock and putting in on the infant's foot. Another similar instance is presented in the next example.

Example 20

VP17 1st visit (04:26–04:27)

In this example, the mother is undressing the infant. In image (a) she lifts the infant's shirt to make room for the infant's head to pass through the shirt opening. In fact, just before this, while pulling the shirt up, she takes a loud inbreathe (see previous example on action announcement).
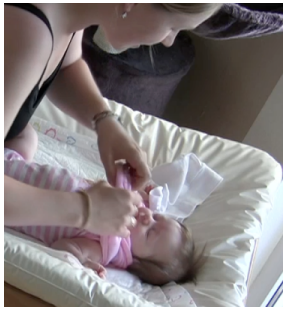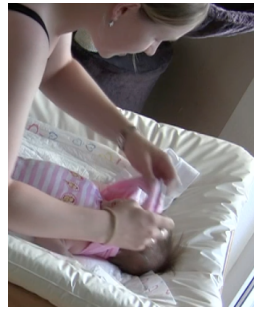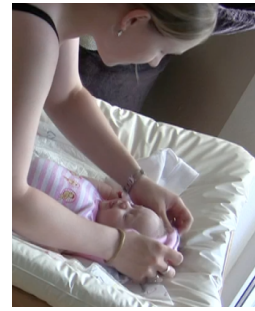
(a) [action preparation]   (b) [action duration]   (c) [action completion]

Figure 6.5.: Use of the hands modality. Throughout the entire duration of her movement the mother is saying *"hui"*.

In image (b) she passes the shirt over the head, and in image (c) pushes shirt down behind the head. While performing this action, she lengthens the vocalization *"hui"* to match the duration of this entire movement.

When playing with their infants, mothers tended to move the infant's body parts synchronously to their speech. Tactile stimulation of this type, that is, in free play interactions, has already been well documented in the literature (Stack, LePage, Hains, & Muir, 1996; Jean, Stack, & Fogel, 2009)[1]. Mothers stimulated infants with touch in synchrony with their speech. This vocal behav-

---

[1] For example, Stack et al. (1996) classify maternal touching into eight possible types of touch:

    a) static touch

    b) stroke/rub/caress/massage

    c) pat/tap

    d) grab/squeeze/pinch

    e) tickle/finger-walk/prod/poke/push

    f) shake/wiggle

    g) pull/lift/flexion/clap, and

    h) other types of touch (kiss, posture change, adjusting clothing, rocking, bouncing, touching with toys)

ior reinforces the body movements and makes them acoustically perceivable. Complementarily, the body of the infant being moved adds a proprioceptive dimension to the vocal signal.

Communicative gestures such as pointing were scarce, and the observation was that the mothers projected their co-expressive hand movements on to the bodies of their infants, moving them to the rhythm of their speech (Zukow-Goldring, 1996; Gogate et al., 2000; Condon & Sander, 1974), and thereby also contributing to the acoustic expression of the movement.

Finally, hands were often used to label actions and objects. For instance, one mother lifted the infant's legs, placing them in the infant's visual field, while simultaneously uttering a demonstrative phrase. A similar example is presented below.

Example 21

VP14 1ˢᵗ visit (13:54–14:23)

This sequence begins with the mother bent over the table and blowing air on the infant's belly (line 1). She looks up towards him, asking where he is. At this moment the infant stretches his arms and touches his mother's mouth (line 2). She responds to this action by kissing the infant's hands, and this opens the topic of interaction, namely the mother starts to talk about the infant's hands.

Transcript 15

```
1 M: ((sound)) (1.8 s) wo bist du? (0.6 s)
   ((sound)) (1.8 s) where are you? (0.6 s)
```

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)
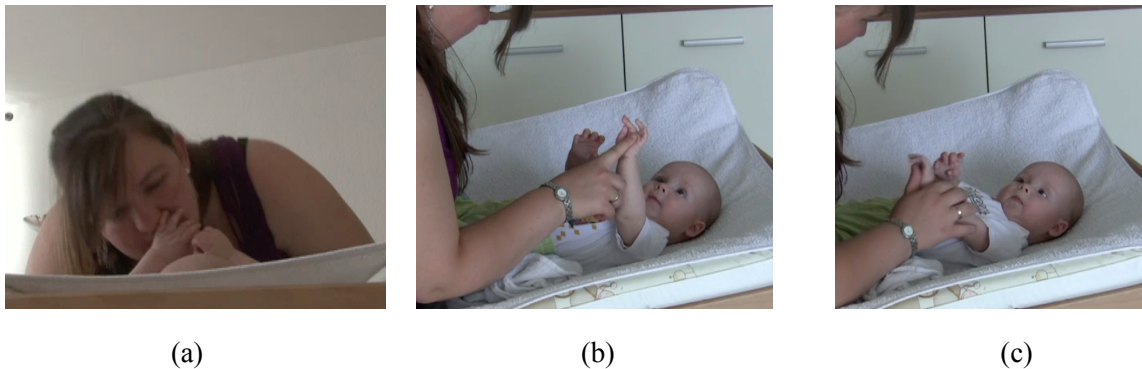
Figure 6.6.: Snapshots referring to transcript line 2

```
2 M: ((inbreathe)) *a((kiss)) deine Finger guck mal *b=
  =da:: *c(1.5 s)
  ((inbreathe)) ((kiss)) your fingers look over=
  the::re (1.5 s)
```

After kissing the infant's hand the mother returns to an upright position. With her pointing finger she lifts he infant's arm showing him his hand (image b) as she utters the deictic *"there"*. She is, thus, touching the referent (the hand) and the infant can feel this part of his body being touched. During the pause, the tension is released and the arm is left to drop (image c). The mother repeats.

```
3 M: eine *dHand (0.5 s) ein Dau*emen ei::ns (0.7 s)
   a hand (0.5 s) a thumb o::ne (0.7 s)
```

In line 3, as she speaks the word *"hand"* she has already lifted the arm, and is holding the infant's hand (image d), here again in perfect synchrony with her speech. She continues to describe the parts of the hand, by showing them to the infant. In image (e), she is separating the thumb from the rest of the hand while saying *"thumb"*. In line 4, the sequence continues with the mother counting all the infant's fingers, while lifting each finger (for the first finger see image (f) and

(d)   (e)

Figure 6.7.: Snapshots referring to transcript line 3

the fourth finger image (g)). In line 5, she continues by defining *"hand"* and *"fist"* (images h-j). The interesting detail here is that she grasps and shakes the infants hand (image h) while saying *"hand"* and *"fist"* in line 5. Then she pauses. While pausing she releases the grip (image i), and accentuates the movement (image j) again, repeating *"fist"* for the second time.

(f)   (g)
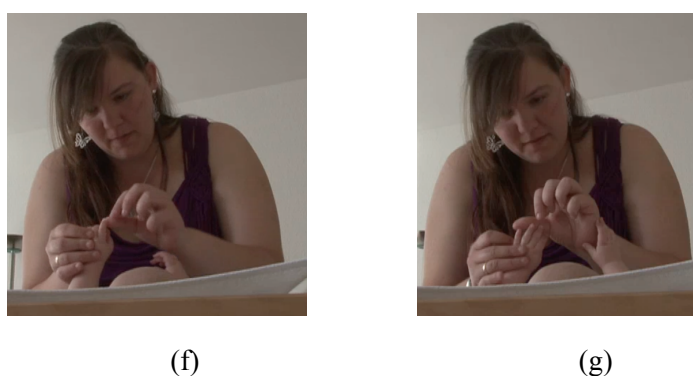
Figure 6.8.: Snapshots referring to transcript line 4

```
4 M: und *feins (- - -) zwei (0.4 s) drei *gvier Finger
   and one (- - -) two (0.4 s) three four fingers
```

<div style="text-align: center;">(h)   (i)   (j)</div>
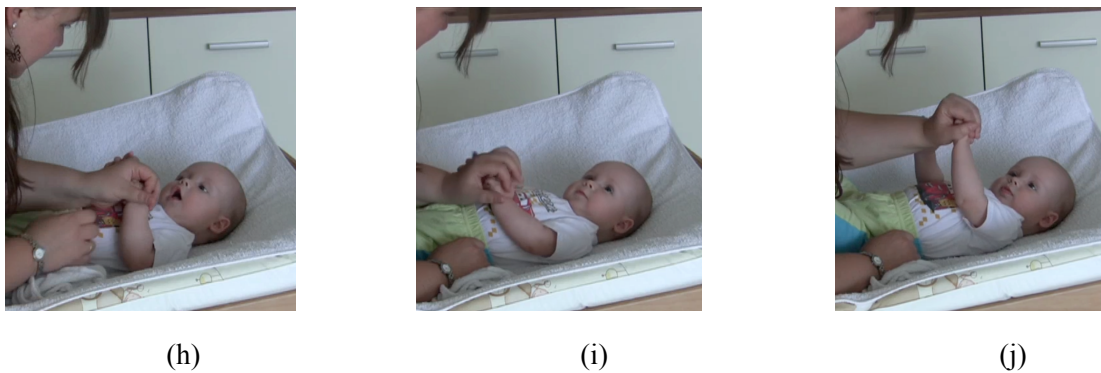
Figure 6.9.: Snapshots referring to transcript line 5

```
5 M: und das ist eine *hganze Hand und da haben wir eine Faust=
    =*i(- - -) ((inbreathe)) eine*j Fau::st (1.1 s)
    and this is an entire hand and there we have a fist=
    =(- - -) ((inbreathe)) a fi::st (1.1 s)
```


<div style="text-align: center;">(k)   (l)</div>
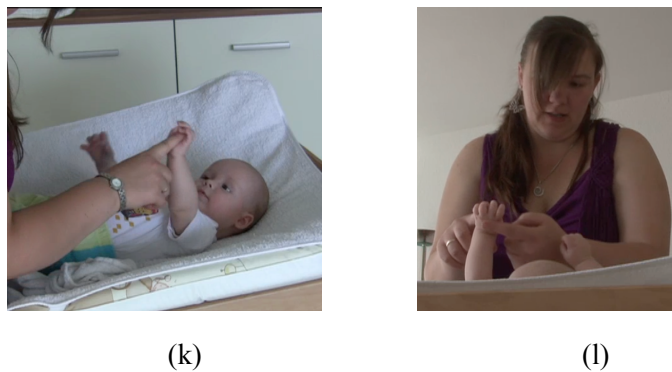
Figure 6.10.: Snapshots referring to transcript line 6

```
6 M: nochmal?*k (0.8 s) ((inbreathe)) ja guck mal *lda=
    =ist die Hand
    once again? (0.8 s) ((inbreathe)) yes look there=
    =is the hand
```

In line 6 she repeats. She lifts the arm again (image k), and this time touches the front side of the hand marking the shape of the hand (image l) with a movement, exactly when she says *"there"*. She continues by touching the arm (line 7, images m-o). The interesting detail here is that she prolongs the word *"arm"* to match the duration of her up-and-down movement on the infant's arm (images m-o).

| (m) | (n) | (o) |
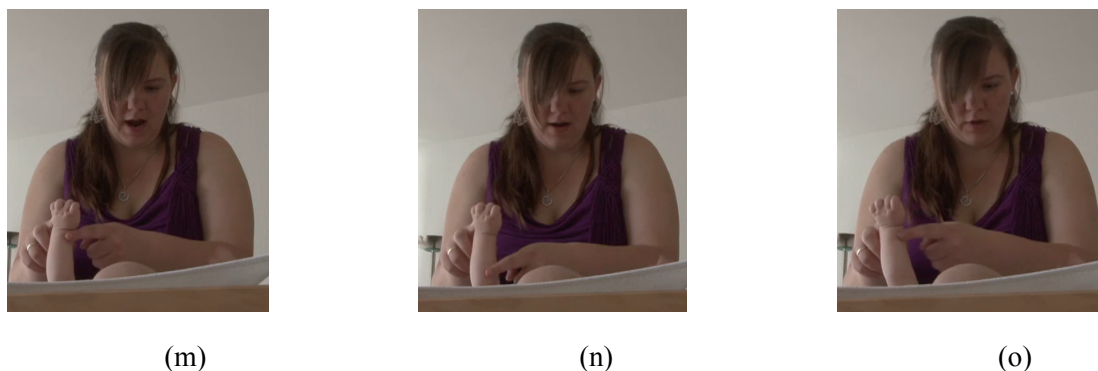
Figure 6.11.: Snapshots referring to transcript line 7

```
7 M: da ist der *m, *n, *oA:rm (0.4 s) [und da] der Oberarm
   there is the a:rm (0.4 s) [and there] the upper arm


8 I: [((vocalization))
```

Further examples include intermodal stimulation by moving the infant's legs fast when saying *"fast"* and slowly when saying *"slowly"*, or stretching the infant's arms above his head while saying *"high"* or *"big"*. In these cases, the mother is communicating the meaning of her utterance, by providing the infant with his or her own bodily experience. This observation is consistent with Zukow-Goldring (1997, p. 219) belief that "[w]ords do not relate arbitrarily to the world as children learn them." Instead, words are lived and experienced in everyday contexts and activities.

6.3.3. Coding class: Upper body

Movement of the upper body was a further modality that was observed to be temporally synchronized with language. Mothers usually moved between a looming or upright position, thus reducing the distance between them and the infant and placing themselves in the focus of the infants' visual field. A closer examination of the vocal context during posture changes revealed that mothers' upper body movement seemed to coincide with phrase boundaries. In such cases, mothers used their body as an additional cue, highlighting where a phrase begins and where a phrase ends. One entire phrase was commonly synchronized to one continuous body movement. Such synchronization was observed more often when it was accompanied by forward movement.

Example 22

VP06 1st visit (00:54–01:00)

This sequence is at the beginning of the interaction. The infant has just woken up, and the mother takes her from bed directly to the changing table for the recording. In line 1, the mother has just taken off the infant's pyjamas and starts by asking the infant how she is doing. Mother and infant are engaging in eye contact, and the infant stretches her arm upwards. The mother reacts to this, by repeating the question and looming forward. In this sequence, this mother utters three questions (lines 2 and 3), and each phrase is accompanied by one body movement: One is leaning forward, the other is bending more forward, the third is returning to the distant position. In this case, it seems that body movement is used to 'package' distinct linguistic events, as a new phrase commences with each new body movement.

Transcript 16

```
1 M: na kleiner Speckbauch? ( - - - ) wie geht's dir?
   so you little potbelly? ( - - - ) how are you?
```

(a) [question start]　　　　(b) [question completed]

Figure 6.12.: Snapshots referring to transcript line 2

```
2 M: *ªwie geht's dir?*ᵇ
```
   how are you?
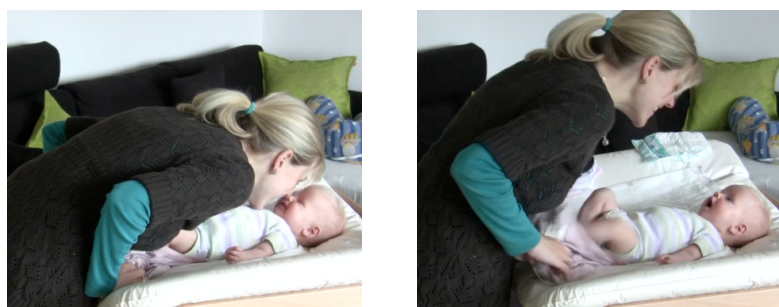
(c) [new question]　　　　(d) [new question]

Figure 6.13.: Snapshots referring to transcript line 3

```
3 M: *ᶜwie geht's dir? (.) gut?*ᵈ
```
   how are you? (.) well?

More specifically, in line 2 the mother starts the forward movement at the beginning of her question (image a) and concludes her question while at the same time stopping over the infant's face (image b). There she pauses shortly and continues with a repetition of the question (line 3)

during which she bends even more forward so that her nose touches the infant's nose (image c). She continues by answering the question with *"good"* but in an interrogative tone, thus asking the infant one more question. As she does this, she returns to a more distant position and stops there (image d). It should also be noted that in the specific example, body movement and vocal behavior do not co-occur perfectly. The movement of the body starts shortly before the onset of speech. Yet, despite this looser temporal overlap, body and speech give the impression of being synchronous, since the upper body does alter its orientation to each new uttered phrase. Thus, in this case described above, the mother marks her vocal behavior multimodally.

In other cases, forward body movement was also employed to highlight certain aspects of a phrase, for example a looming movement at the end of a question, as if the mother were signaling that it was the infant's turn to answer. In this case, the synchronized body movement seems to be used in a similar way to a discourse marker, handing over the turn to the infant.

The upper body movement was also embedded in games. Infants were excited when mothers moved between a forward and upright position. Such movements cycled repetitively and were accompanied by sounds, interjections, songs, or repeated phrases.

Example 23

This example illustrates such a game. The mother moves forward approaching the infant, and what follows is a rhythmical approach-withdrawal game in which the mother keeps repeating the same word, as she makes nose-to-nose contact with the infant. The infant responds to this with excitement and starts to laugh. This probably motivates the mother to continue the game. Concerning the synchrony between the mother's speech and the movement of her upper body, this sequence is an example of the manner in which body movement segments various hierarchically different units of the mother's utterance: from phrase to word to syllable.

VP09 4th visit (01:23–01:40)

Transcript 17

The sequence begins with the mother leaning forward towards the infant while commenting

on her action, namely having fixed the diaper in the back and folded it (line 1). This phrase she utters in synchrony with her forward movement. In image (a), she starts her phrase, and in image (b) she stops to take an inbreathe to continue her phrase. This is similar with the previous example with the synchronization of the upper body with utterance boundaries.

(a)   (b)
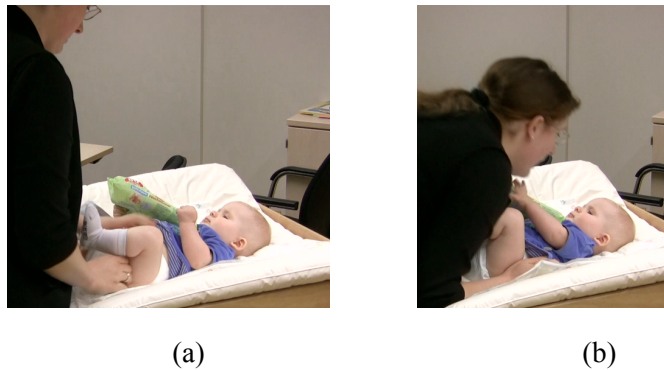
Figure 6.14.: Snapshots referring to transcript line 1

```
1 M: *aHinten dran und einschlagen *b((inbreathe)) dann machen=
=wir den Body *czu*d (0.7 s)
  fix on the back and fold it ((inbreathe)) then we=
  =close the onesie (0.7 s)
```


(c)   (d)

Figure 6.15.: Snapshots referring to transcript line 1

After the inbreathe the mother continues to describe her next action which will be buttoning the infant's onesie. Here, she synchronizes her body movement with the boundaries of a single word *"zu"*. In image (c) her body is still close to the infant and in (d) just after speaking this word she has moved away to a more upright position. This phrase then becomes the game. She goes on repeating this phrase in a rhythmical pattern, while also repeating the movement pattern illustrated in images (e-g): she moves forward to make nose-to-nose contact at the word *"Body"* (image f) and moves back in synchrony with *"zu"*. More specifically, the movement is so well timed so as to touch the infant's nose exactly at the stressed syllable *"Bo"* in *"Body"*, thus marking an even smaller segment of a syllable within her phrase.

| (e) | (f) | (g) |
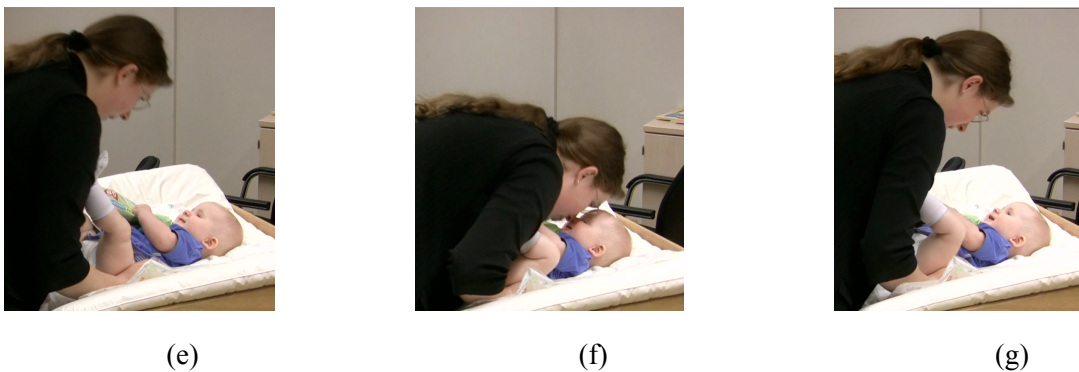
Figure 6.16.: Snapshots referring to transcript line 2

```
2 M: dann *e machen wir den Bo*fdy zu *g (0.7 s)
  then we close the onesie (0.7 s)
3 I: ((laughing))
4 M : [machen wir den Body zu? (0.9 s)
  do we close the onesie? (0.9 s)
5 I : [((laughing))
```

The infant reacts to this with loud laughter, and thus, the game is established. In lines 2 and 4 the mother repeats the same pattern. In line 6, there is a break, or variation, in the pattern. The

mother comments on the infant's laughter. In doing so, she looms in and makes nose-to-nose contact (image h), and moving her head three times to the left and right she rubs her nose on the infant's nose in synchrony with the three words in her phrase *"findest"*, *"du"*, and *"das"*. Here again a segmentation of the words in her phrase occurs, and the infant can feel this patterning through the tactile sensation on her nose. Also, the mother marks the end of the phrase, by moving back to a more upright position in synchrony with the phrase boundary (image i).

(h)   (i)
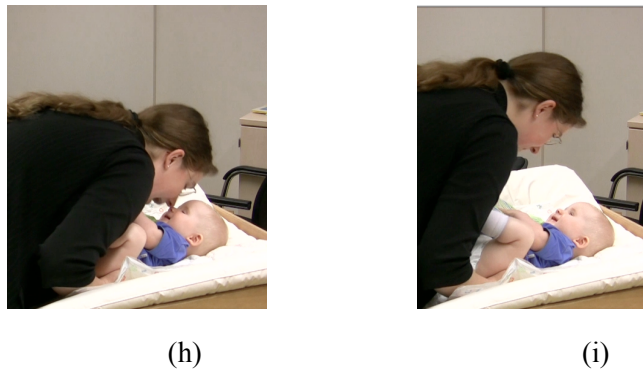
Figure 6.17.: Snapshots referring to transcript line 6

```
6 M: ja::*h [findest du das lustig?*i (1.1 s)
   ye::s do you find this funny (1.1 s)
7 I: [((laughing))
```

Line 6, illustrates the employment of multiple modalities. The modalities are combined to mark differently sized units of the mother's utterance. The torso marks phrase boundaries, whereas the head and the nose-to-nose tactile sensation mark the boundaries of the words within the utterance. This nesting of modalities within modalities is illustrated schematically in Figure 6.18. This overlap of multiple modalities will be discussed later in this chapter, separately.

In line 8, the theme of the game changes slightly in that the mother changes what she says when looming in to make nose-to-nose contact. Here, again, she synchronizes her movement so as to touch the infant's nose exactly after the syllable *"na"* in *"nasenstüber"* (image j), which, apart

Figure 6.18.: Schematic representation of the nested modalities marking hierarchically different units of an utterance. The torso movement (green) marks the boundaries of the utterance; the head movement (grey) marks subunits, i.e. words within it. The figure refers to transcript 17, line 6.

from forming a temporal repetitive pattern, ensures that the infant feels her nose being touched at exactly the moment in which the mother says a word containing the word *"nose"*. This pattern is identically repeated in lines 10 and 12.

(j)   (k)
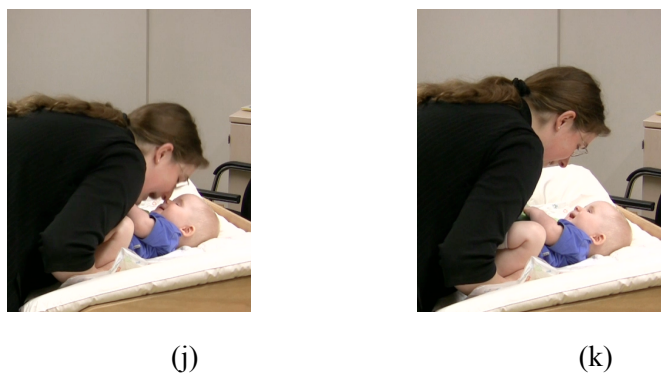
Figure 6.19.: Snapshots referring to transcript line 8

```
8 M : na*ʲsenstüber*ᵏ (0.7 s)
```

```
   tap on the nose (0.7 s)
9 I: ((laughing))
10 M: nasen[stüber (0.7 s)
   tap on the nose (0.7 s)
11 I : [((laughing))
12 M : nasenstüber (0.7 s) hä?
   tap on the nose (0.7 s) huh?
```

This example further illustrates the temporal patterning of the games in which synchrony is used. More specifically, the rhythmical aspect of this sequence is visible both in the repetition of the phrases by the mother, but also in the duration of the pauses after her phrases. As can be seen in lines 1, 2, 8, 10, and 12, the pauses are identical. Yet, within this regularity, there is also variability. The movement of the body and the timing remain the same, but the mothers' words change. Lines 1, 2 and 4 have one theme, and lines 8, 10 and 12 another. Stern and Gibbon (1979) have argued that variation in maternal behavior maintains infant attention, because it violates the infant's expectancies. In this case, this mother uses variation in the theme of the game, thereby making her playful behavior more unexpected to the infant. This practice was also discussed in the previous chapter describing the mothers' practices in maintaining attention.

6.3.4. Coding class: Head

A large majority of the mothers made extensive use of head movements. These included engaging in an accentuated shaking of the head from side to side or nodding in synchrony with their speech. They made large, exaggerated movements with their heads when saying *"yes"* (up-down movement) or *"no"* (left-right movement), as illustrated in the following examples.

Example 24

VP08 4th visit (05:07–05:08)

(a)   (b)   (c)

Figure 6.20.: Example 24:Mother performing up-down head movement while saying *"yes"*. The vowel in *"ja"* is prolonged so as to accommodate the down-up movement.

Example 25

VP08 1st visit (00:29–00:30)

(a)   (b)   (c)

Figure 6.21.: Example 25: Mother performing left-right head movement while saying *"no"*. The diphthong in *"nein"* is prolonged so as to accommodate the left-right movement.

Another observation was that some mothers also moved their heads in temporal synchrony with their speech, for instance, changing the direction of their head at the onset of each word, or sometimes even syllable in a phrase, as if reinforcing the vocal signal by rendering word boundaries more salient (see example 23 above).

Head movements were also used to highlight a specific word in the phrase. For example, one mother repeated a sentence twice, and both times she moved her head in the same way when pronouncing the same word (see example below).

Example 26

VP08 1ˢᵗ visit (06:53–06:58)

Transcript 18

(a)   (b)   (c)
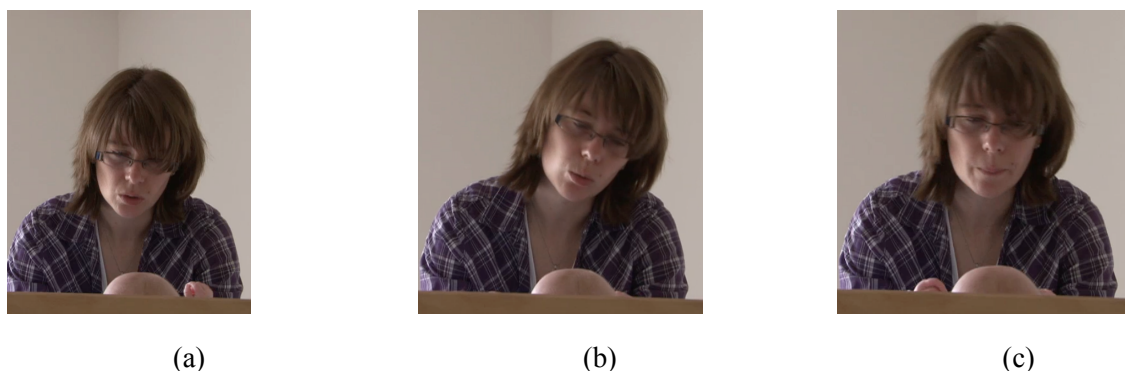
Figure 6.22.: Snapshots referring to transcript line 1

```
1 M: *ªnaja nun nimm*ᵇ es*ᶜ
  well don't take it
2 I: ((vocalization))
```

In this example, the mother used the head movement to stress the word *"nimm"*, making it stand out from the other words in her phrase. In line 1, the head is in a starting position (image a) at the beginning of her utterance. In image (b), she has just spoken the word *"nimm"* and returns to

the starting position in image (c). In this case the mother also stresses this word in her speech, as she uses contrastive prosody. She, thus, reinforces her message by incorporating both prosody and head movement. The same happens also in line 3 illustrated in images (d-f).

(d)   (e)   (f)
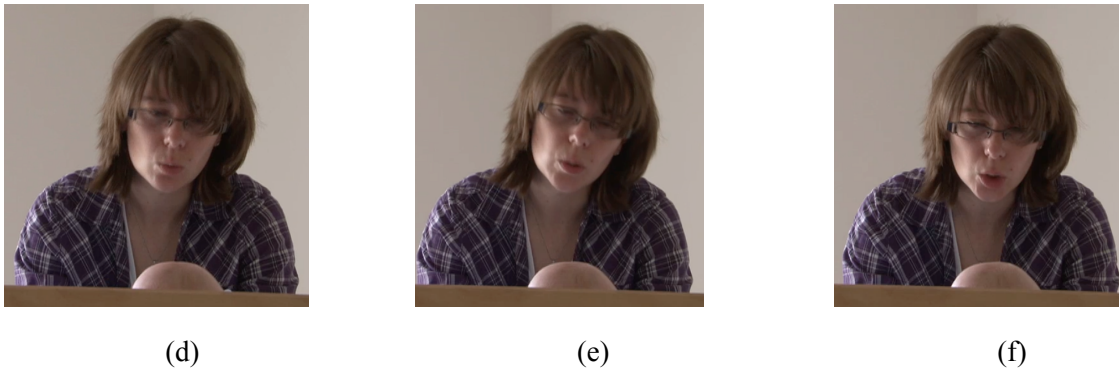
Figure 6.23.: Snapshots referring to transcript line 3

```
3 M: *dnun nimm*e es mir mal nicht *fkrumm
     well don't take it the wrong way
```

Here again, body movement is used to highlight a certain aspect of language, making it stand out in comparison to the rest. Furthermore, interpreting this example from a more macroscopic perspective, it becomes evident that the mother makes use of a different type of temporal overlap, namely, one during which one modality (in this case the vocal) is activated continuously, whereas the second modality (the head) is added at specific points in time to highlight one specific element. Many similar multimodal cases were observed in the data. Mothers also used the half-nod head movement (movement in one direction either upward or downward) while encouraging their infant, as illustrated in the following example.

Example 27

VP03 1st visit (00:26–00:28)

Transcript 19

(a)   (b) [introducing question]   (c) [question]

Figure 6.24.: Snapshots referring to transcript line 1

1 M: *a((inbreathe)) *b geht's dir gut*c?

6.3.5. Coding class: Face

Example 28

Changes in facial expressions were systematically observed together with certain behaviors and in temporal co-occurrence with language. It should be noted, however, that there was great variability between subjects; some mothers made extensive use of this modality, whereas others exhibited a limited used of vocal–face synchrony.

As in the case of the head and torso modality, changes in facial expressions, especially an exaggerated rising of the eyebrows, highlighted or framed specific parts of an utterance or phrase. On some occasions, mothers raised their eyebrows at the onset or the end of a phrase. This practice may serve to emphasize phrase boundaries, indicating to the infant when speech begins, and at what point it is going to end.

VP10 1st visit (07:40–07:43)

(a) [highlight prepa- (b) [builds up high- (c) [concludes high- (d) [concludes utter-
ration] light] light] ance]
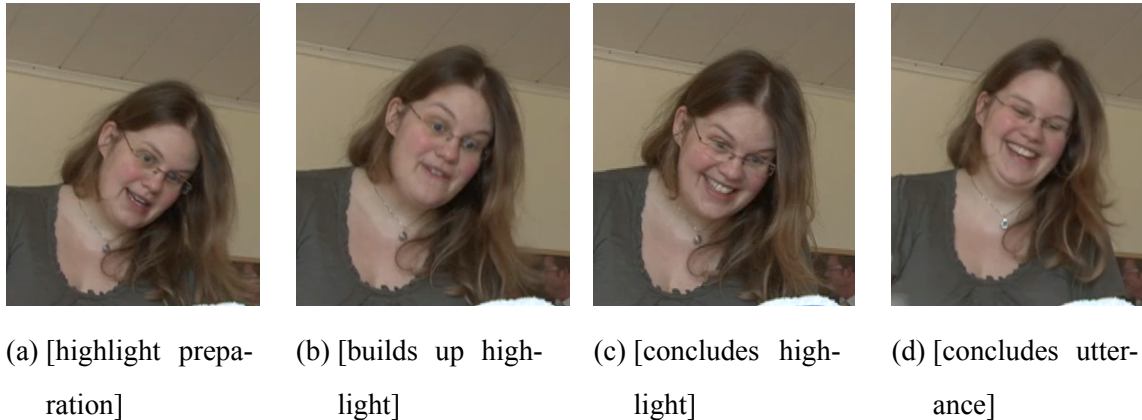
Figure 6.25.: The mother raises her eyebrows to highlight her utterance.

The sequence starts with the mother's face in neutral position (image a), during which she calls the infant name. In (image b) she raises her eyebrows as she starts to utter the *"f"* in her phrase *"Fertig"* (eng. = finished). In image (c) she starts to lower her eyebrows as she has just uttered her phrase and in image (d) the highlight is no longer active.

Example 29

Use of the surprised face and raised eyebrows was generally observed as an attention-getter. Often, facial movement occurred in combination with emotionalization sounds (loud inhaling or panting) and with flicking or smacking sounds produced by the mother when trying to attract or maintain the infant's attention.

VP05 1st visit (00:21–00:22)

    (a)   (b)   (c)
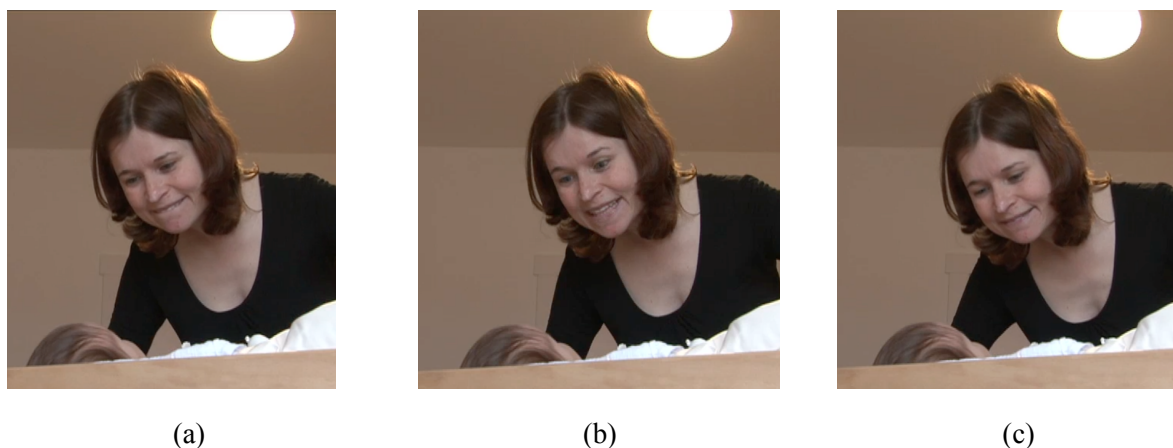
Figure 6.26.: The mother raises her eyebrows during a click sound.

6.3.6. Coding class: Lips

Within the coding class face, lip movements occurred as an interesting case of making transparent what the vocal activity is referring to. Reference is a problem, because, usually, it is difficult for an infant to make a link between a label and an object that is labeled (see a description of this problem in e.g., (Rolf et al., 2009)). As suggested in the introduction, this link can be facilitated by attention-directing gestures (Gogate & Bahrick, 2001; Gogate et al., 2009). In the case of kissing, the child can experience it visually (seeing the lips moving and making a sound) as well as tactilely (being kissed and hearing the sound). Thus, the reference of the lips making a kiss to the action of kissing is straightforward.

In the data, mothers used kissing as part of playful behavior. For example they repeatedly kissed the infant on the cheeks alternating from left to right, while accompanying the act of kissing with their voice. Mothers also introduced an oncoming kiss, by starting to vocalize while approaching to kiss the infant. This way, the communicative intention was introduced by the voice, and the stroke (the kiss) could be felt tactilely. This is a good example illustrating that the

179

synchrony of speech and action goes beyond cases in which the modalities occur simultaneously in time (Zukow-Goldring & Rader, 2005). In this case, the voice actually introduces the action. This shall be discussed below.

6.3.7. Summary of qualitative results

So far, qualitative observations from all identified classes of body movement have been presented. Also, the variety of spontaneous ways in which, verbal behavior co-occurring with body movement was used communicatively to structure actions, has been illustrated. In sum, it has been shown how language introduces new actions, breaks action sequences into smaller chunks, highlights the duration of actions, or marks the goals, or ends, of actions. At the same time, examples have been provided supporting the hypothesis that action in the form of body movements (as classified in Table 6.1), when co-occurring with language, renders language visible and perceivable to the infants by coupling it to bodily experience. It was shown how body action is employed to mark boundaries and segment speech, but also to underline and highlight certain aspects of speech. Finally, it was shown how action is used to convey the meaning of an utterance. Yet most importantly, it has been shown that when putting the coded overlapping behavior into its situational context, various global patterns of temporal overlap emerge, which are referred to as 'types of synchrony'. Inspired by the basic interval relations (Nebel & Bürckert, 1995; Kopp & Bergmann, 2010), it was observed that in the data, two intervals – in this case, the vocal and motor modalities – overlap in various ways and form temporal relationships that are summarized in the Figure below.

For example, the vocal signal can introduce or announce the action, that is, occur earlier than the actual action (type b). This type might help sustain infant attention, because the interval during which one or two modalities are active is prolonged though the offsetting of modality onset. Language can also finish or comment on the action performed, that is, occur subsequent to the referred action. Type (c) could be used to verbally mark the beginning of an action, communicating to the infant that something new is being initiated. Language use can also partially
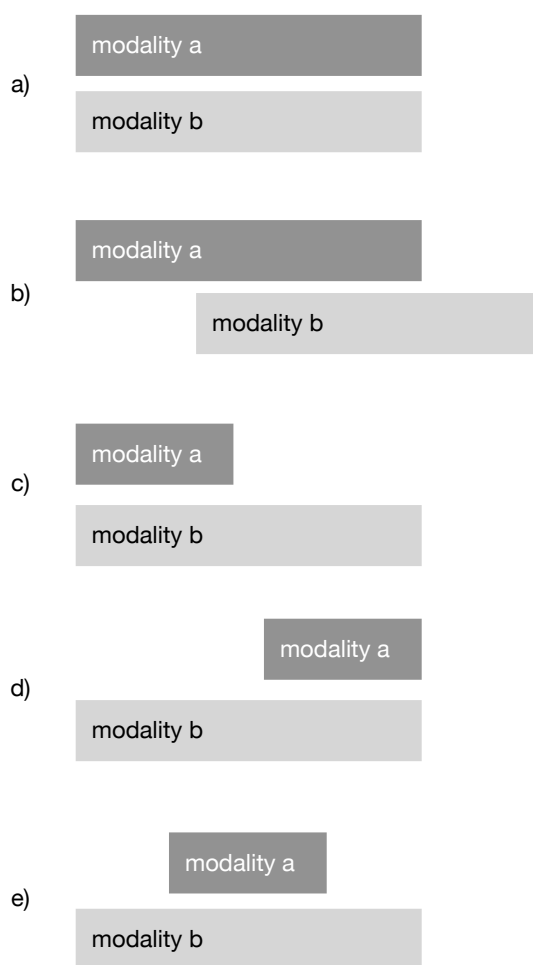
Figure 6.27.: Temporal relations between speech and movement.

overlap when an action's goal is marked (type d). Complementarily, in type (d), an action might be used to highlight the end of an utterance, signifying to the infant that it is his or her turn to speak. Finally, type (e) might provide a longer interaction frame during which one element is specially highlighted.

6.4. Results from quantitative analysis

This section includes:

a) the descriptive statistics of the coded behavioral classes over the entire interaction time

b) descriptive statistics for the use of body classes over the time during which the mother vocalized, and

c) a report on the co-occurrence of multiple body classes as observed in the data.

In the different subsections the data are first presented for the three-month-old infants and then for the six-month-old infants. Finally, at the end of the section a comparison is made between the data of these two visits.

6.4.1. Synchrony between language and action over interaction time

Three-month-old infants

For each modality, the percentage of the durations over the interaction time were calculated and subsequently averaged over all subjects. For the three-month-olds, vocal-hands was the modality used predominantly by mothers during the interaction (16.97 % of the time ; $SD = 7.53$). The second most frequently employed modality was vocal-head (7.99 % of the time ; $SD = 5.01$). In third place the vocal-facial modality (5.68% of the time ; $SD = 4.07$) was observed. Finally, upper body movement was synchronized with vocal behavior for an average of 3.90% ($SD = 2.24$) of the overall interaction time, and the lips body class took up 0.97% ($SD = 1.25$). The complete data can be found in Appendix Table B.1.

Next, the single durations of synchrony events were extracted and subsequently averaged over all subjects. This gave an impression of how fast or slow the pace is, in which the synchrony is provided. The longest events occurred when using the hands modality (1.47 sec ; $SD = 0.36$) , followed by the face (1.13 sec ; $SD = 0.43$), the head (1.05 sec ; $SD = 0.33$), the torso (1.13 sec

; *SD* = 0.28) and last the lips (0.86 sec ; *SD* = 0.88). The data can be found in Appendix Table B.2.

Six-month-old infants

For the six-month-olds, the vocal-hands modality was used predominantly (15.54 % ; *SD* = 7.89). The second most frequently employed modality changed and became the vocal-facial (10.84 % ; *SD* = 8.32). The next most used modality was the upper body movement (9.27 % ; *SD* = 5.84), and the vocal-head modality came in fourth place (5.18 % ; *SD* = 4.90). Finally, the lips body class took up 0.91% (*SD* = 0.83). The complete data can be found in Appendix Table B.4.

As for the average duration of synchrony events, the intervals for the hands modality were the longest (0.94 sec ; *SD* = 0.29), followed by the facial modality (0.87 sec ; *SD* = 0.29=, the lips (0.82 sec ; *SD* = 0.96) and then the torso movements (0.73 sec ; *SD* = 0.25). The shortest intervals were for the head modality (0.52 sec ; *SD* = 0.23). The full data can be found in Appendix Table B.5.

6.4.2. Overall synchrony between action and language over interaction time

The results presented above show the extent to which the mothers in the data used each one of the synchrony modalities. Since these coding classes are not mutually exclusive, that is, the modalities can be used simultaneously at a given point in time, they are not additive. To find out the total proportion of the interaction time any (or all) modality was active, the duration of the *overall synchrony* was calculated. This is schematically illustrated in Figure 6.28.

Three-month-old infants

In Appendix Table B.3 the results are listed for the percentage of overall synchrony over inter-action time for the three-month-old infants. Mothers synchronized their body movements with their verbal behavior during 29.1% (*SD* = 10.1) of the interaction time.
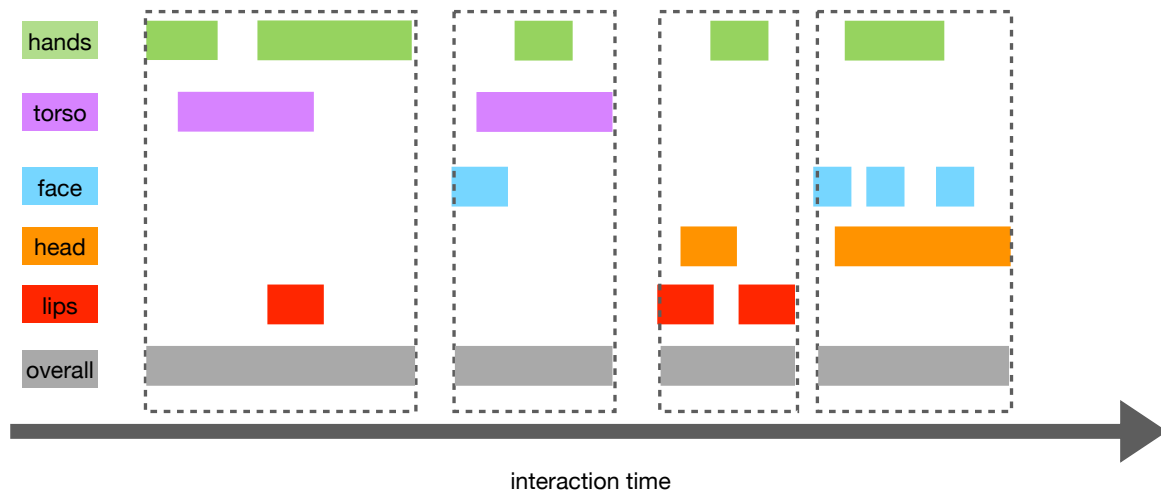
Figure 6.28.: Schematic representation of overall synchrony. The intervals marked in grey represent the union of the intervals coded for each modality.

Six-month-old infants

For the six-month-olds Appendix Table B.6 reports the results. Mothers synchronized their body movements to their language during 27.06% ($SD = 12.15$) of the interaction time.

6.4.3. Comparison synchrony over interaction time at three vs. six months of infants' age

To test whether age influences the occurrence of multimodal practices, a repeated-measures ANOVA was calculated. This revealed no main effect for AGE, *F(1, 16) = 2.03, p = .174, η^2 = .11*, but an interaction effect for AGE x SYNCHRONY, *F(4, 13) = 18.77, p < .001, η^2 = .852*, suggesting that the overall use of synchrony did not change with infants' age, but age affected the synchrony modalities differentially. Post-hoc pairwise comparisons are reported in Appendix Table B.7.

Figure 6.29.: Comparison of the mean percentage of interaction time for the various synchrony modalities at the different points of observation. ***p < .001. **p < .01. Paired sample tests (2-tailed).

As illustrated in Figure 6.29, the overall use of maternal coordinated speech and body movement did not change significantly between the two observation points. However, which modalities convey the synchrony seems to be age-dependent. More specifically, in comparison to interacting with three-month-olds, when interacting with six-month-olds, mothers (a) significantly increased the use of the facial modality ($t(16) = 3.54, p = .003$) while significantly reducing the use of the head modality ($t(16) = 3.18, p = .006$), and (b) significantly increased movements of their upper body (looming in and out) ($t(16) = 4.76, p < .001$). The use of hands synchrony and the lips modality was not influenced by the infant's age.
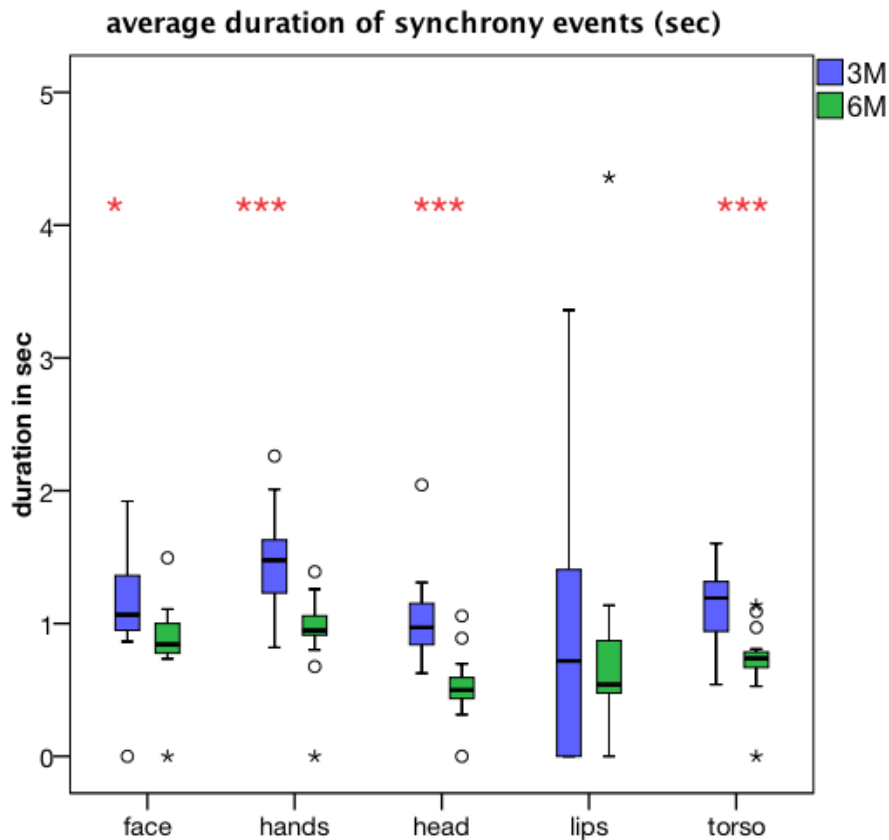
Figure 6.30.: Comparison of the average duration of synchrony events in seconds in the various synchrony modalities at the two points of observation.***p < .001. *p < .05. Paired sample tests (2-tailed).

To investigate whether the duration of synchrony events is influenced by infants' age, a repeated-measures ANOVA was calculated. This revealed a main effect for AGE, $F(1, 16) = 23.02$, $p < .001$, $\eta^2 = .59$, and an interaction effect for AGE x SYNCHRONY DURATION, $F(4, 13) = 3.63$, $p < .03$, $\eta^2 = .53$, suggesting that age influenced the duration of synchrony events and affected all modalities. Post-hoc pairwise comparisons are reported in Appendix Table B.8.

As illustrated in Figure 6.30, the duration of the synchrony events became significantly

shorter in the interactions with the six-month-olds for all modalities except for the lips: for the FACE: (*t(16)* = 2.26, *p* = .038) ; HANDS: (*t(16)* = 7.32, *p* < .001) ; HEAD: (*t(16)* = 6.69, *p* < .001) ; TORSO: (*t(16)* = 5.36, *p* < .001).

6.4.4. Synchrony between language and action over spoken time

The above section revealed the proportion of the interaction during which the mothers provided multimodal behavior. This is a global measure of maternal synchrony. To find out to which extent mothers reinforce their vocal behavior with body movements, the synchrony was considered in relation to this vocal behavior. For this, the duration of maternal vocal behavior was annotated with PRAAT and revealed the duration of the overall spoken interaction. The durations for both visits are listed in Table 6.2.

Table 6.2.: Spoken time at three and six months

Infant Age	N	Mean	*SD*	Minimum	Maximum
3M	17	39.38	10.85	23.55	56.23
6M	17	36.14	13.19	0.80	53.53

When the infants were three months old the mothers vocalized 39.38% (*SD* = 10.85) of the time they were interacting with their infants. When the infants were six months old they vocalized 36.14% (*SD* = 13.19) of the interaction time. A paired sample *t*-test computed showed no significant difference in the proportion of the interaction mothers spent vocalizing (*t(16)* = 1.58, *p* = .13).

The next question was how much of this vocal behavior is synchronized with body movement. For this, the percentages of coded classes of body movement over the total spoken time were calculated .

Three-month-old infants

For the three-month-old infants, 42.80 % (*SD* = 13.92) of the vocal stimulation that mothers provided was synchronous with their hand movements, 19.09 % (*SD* = 9.43) with movements of their head, 14.08 % (*SD* = 9.05) with their facial expressions, 10.30% (*SD* = 6.18) with movement of their upper body, and 2.60 % (*SD* = 3.65) of vocal behavior consisted of the sounds produced by kissing the infant. The complete data can be found in Appendix Table B.9.

Six-month-old infants

For the six-month-olds, 40.46 % (*SD* = 17.87) of the vocal stimulation that mothers provided was synchronous with their hand movements. Here, the next most frequent modality changed to the facial expressions 27.23 % (*SD* = 17.50) followed by 24.28 % (*SD* = 14.73) of vocal behavior being coordinated with movement of the upper body, and 12.37 % (*SD* = 9.32) with movement of the head. Finally, 2.44 % (*SD* = 2.15) of vocal behavior consisted of the sounds produced by kissing the infant. The complete results can be seen in Appendix Table B.11.

6.4.5. Overall synchrony between action and language over spoken time

As in the previous section on synchrony over interaction time, the coding classes are not mutually exclusive, that is, the modalities can be used simultaneously at a given point in time; they are not additive. For this reason, the total proportion of the spoken time during which any (or all) modality was active was calculated.

Three-month-old infants

For the three-month-olds 73.34 % (*SD* = 12.71) of all vocal stimulation was performed multi-modally (as can be seen in Appendix Table B.10), that is, it overlapped with body movements.

Six-month-old infants

When the infants were six months old 69.73% (*SD* = 22.92) of the vocal stimulation was per-formed multimodally (Appendix Table B.12).

6.4.6. Comparison of synchrony over spoken time provided in the different age groups

To test whether age influences the extent to which maternal behavior is reinforced by multimodal practices, a repeated-measures ANOVA was calculated. This revealed no main effect for AGE, $F(1, 16) = 4.22$, $p = .057$, $\eta^2 = .21$. The main effect was marginally significant. Furthermore, it revealed an interaction effect for AGE x SPOKEN SYNCHRONY, $F(4, 13) = 16.67$, $p < .001$, $\eta^2 = .837$, suggesting again that the overall use of synchrony did not change with infants' age, but age affected the synchrony modalities differentially. Pos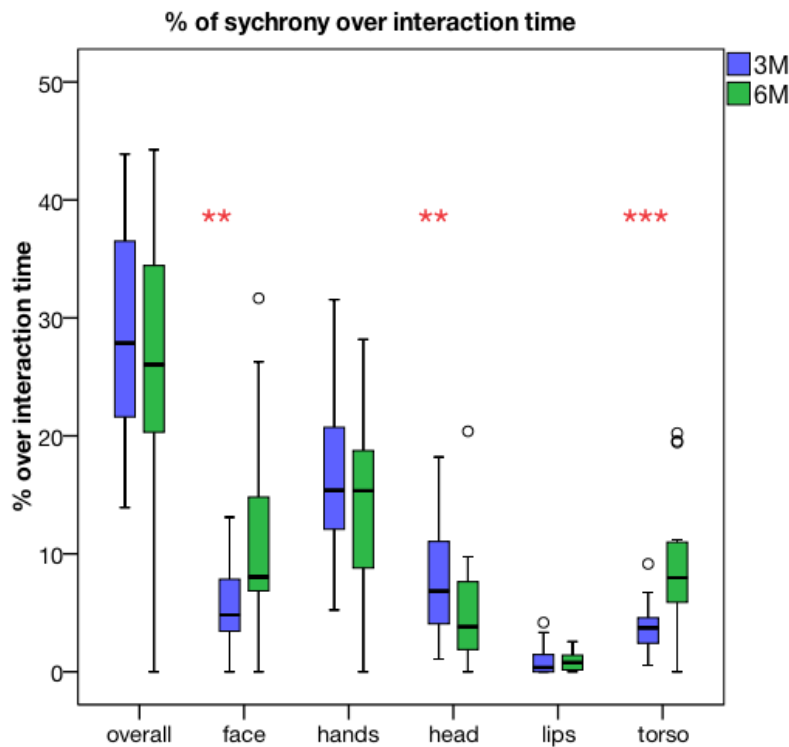t-hoc pairwise comparisons are reported in Appendix Table B.13. As illustrated in Figure 6.31, the overall use of maternal coordinated speech and body movement did not change significantly between the two observation times. What changed was which modality was used to accompany the mothers' vocal behavior. In comparison to the vocal behavior to the three-month-olds, when vocalizing to their six-month-olds, mothers (a) significantly increased the use of the facial modality ($t(16) = 4.09$; $p = .001$) while significantly reducing the use of the head modality ($t(16) = 2.92$; $p = .010$), and (b) significantly increased movements of their upper body (looming in and out) ($t(16) = 5.31$; $p < .001$). The use of hands synchrony and the lips modality was not influenced by the infants' age.
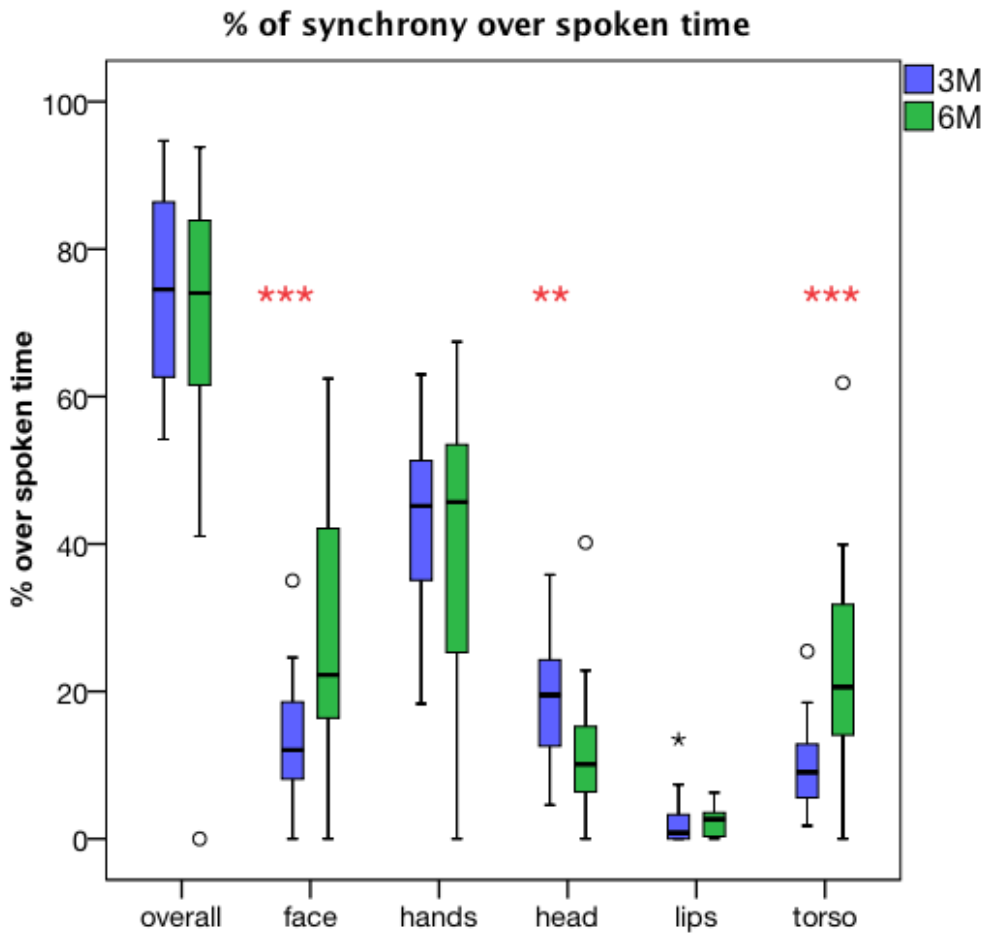
Figure 6.31.: Comparison of the mean percentage of the spoken time for the various synchrony modalities at the two observation points.***p < .001. **p < .01. Paired sample tests (2-tailed).

6.4.7. Co-occurrence of multiple body classes

As already discussed in previous sections, the coding classes identified in the present analysis are not mutually exclusive; more than one modality could be employed simultaneously. The

general qualitative observation was that this seemed to occur often in the data. More specifically, in example 23 it was shown how multiple modalities are combined so as to mark different sized units of the mother's utterance. The torso marks phrase boundaries, whereas the head and nose-to-nose tactile sensation marks the boundaries of the words within the utterance. This nesting of modalities within modalities was also illustrated schematically in Figure 6.18.

This qualitative observation led to the systematic analysis of the co-occurrence of multiple modalities. This was achieved using an INTERACT®built-in statistical analysis module. All possible combinations of modalities were identified, and the percentages of durations over spoken time were calculated for sets of two, three, four, and all modalities together.

Combination of two modalities and vocal behavior

For two modalities there were 10 possible combinations. The descriptives for these combinations can be seen in Appendix Table B.14 and Table B.15 for the three-month-old and six-month-old infants, respectively.

Three-month-old infants

Most predominantly, FACE and HEAD co-occurred for an average of 7.26 % ($SD = 4.89$) of spoken time, followed by HANDS and TORSO (4.01%; $SD = 3.36$), and HANDS and HEAD (3.64 % ; $SD = 3.54$) as the third most frequent pair. Fourth in the row was the pair HANDS and FACE, followed by TORSO and HEAD, and FACE and TORSO. The remaining modality pairs occurred less than 1% of the spoken time. The frequency of occurrence of the co-occurring modality pairs is illustrated in Figure 6.32.

Six-month-old infants

With the six-month-old infants there were differences in the employment of overlapping modality pairs. Most predominantly used were the HANDS and TORSO modalities–which were the second most frequent combination for the three-month-olds–occurring for an average of 9.89

Combination of 2 modalities at 3 months
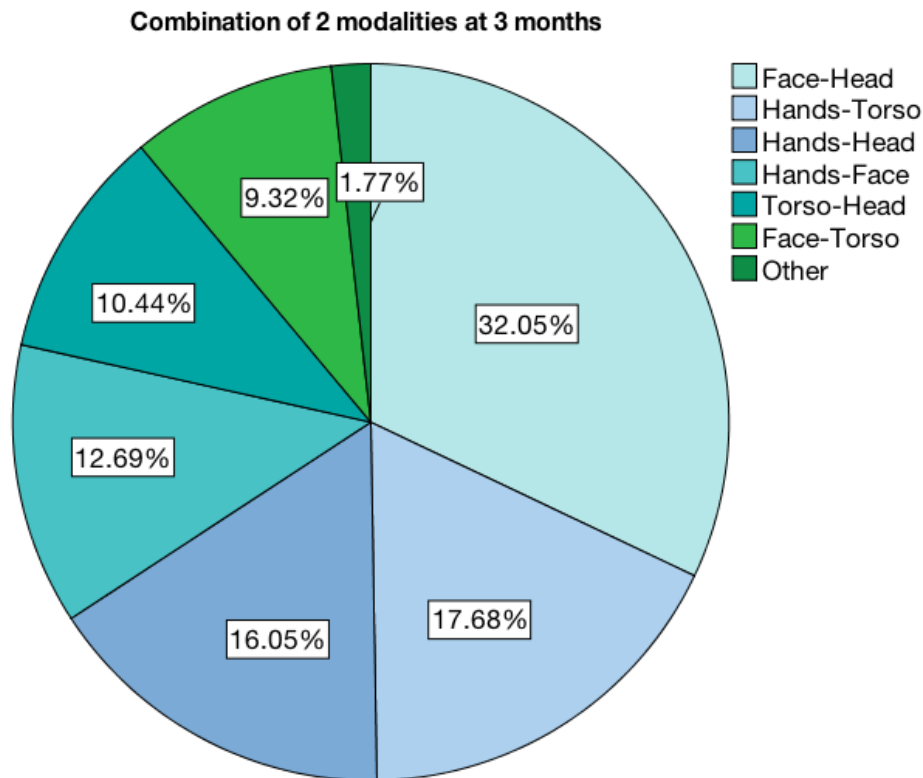
Figure 6.32.: Frequency of co-occurrence of two modalities at three months of age

% (*SD* = 5.97) of spoken time. This was followed by HANDS and FACE (9.55%; *SD* = 7.40)–which were the fourth most frequent pair for the three-month-olds, and FACE and TORSO (9.94 % ; *SD* = 7.15)–which was the sixth most frequent pair for the three-month-olds. The fourth most frequent modality pair was FACE and HEAD (7.30% (*SD* = 7.22), which was the most frequent pair for the three-month-olds. This was followed by TORSO and HEAD (4.33% (*SD* = 5.78) and HANDS and HEAD (3.01% (*SD* = 2.61). Here again, the remaining modality pairs occurred less than 1% of the spoken time. The frequency of occurrence of the co-occurring modality pairs is illustrated in Figure 6.33.

Combination of 2 modalities at 6 months
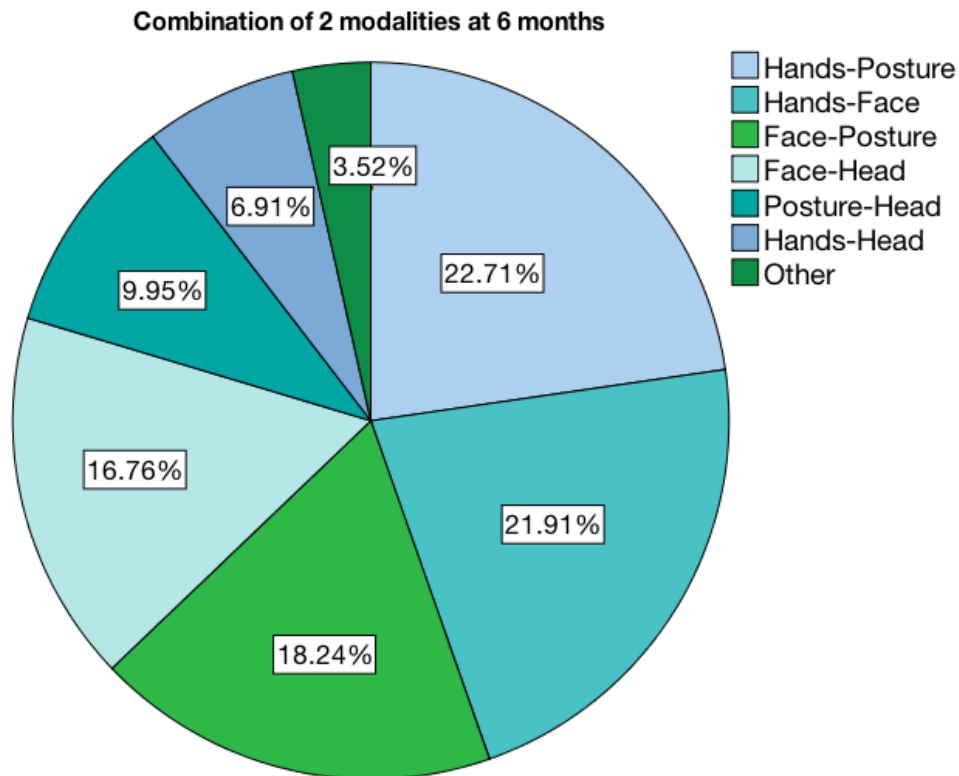
Figure 6.33.: Frequency of co-occurrence of two modalities at six months of age

Combination of three modalities and vocal behavior

For sets of three modalities there were 10 possible combinations. The descriptives for these combinations can be seen in Appendix Table B.16 and Table B.17 for the three-month-old and six-month-old infants, respectively.

Three-month-old infants

When considering groups of three overlapping modalities, FACE-TORSO-HEAD (1.03 % of spoken time; $SD = 1.13$) as well as HANDS-FACE-HEAD (0.99 % ; $SD = 1.24$) were combined most frequently. These were followed by HANDS-FACE-TORSO (0.8 % ; $SD = 1.33$) and

HANDS-TORSO-HEAD (0.68% ; *SD* = 1.13). The remaining modality combinations occurred less than 0.1% of the spoken time or not at all. The frequency of occurrence of the co-occurring modality pairs is illustrated in Figure 6.34.

Combination of 3 modalities at 3 months
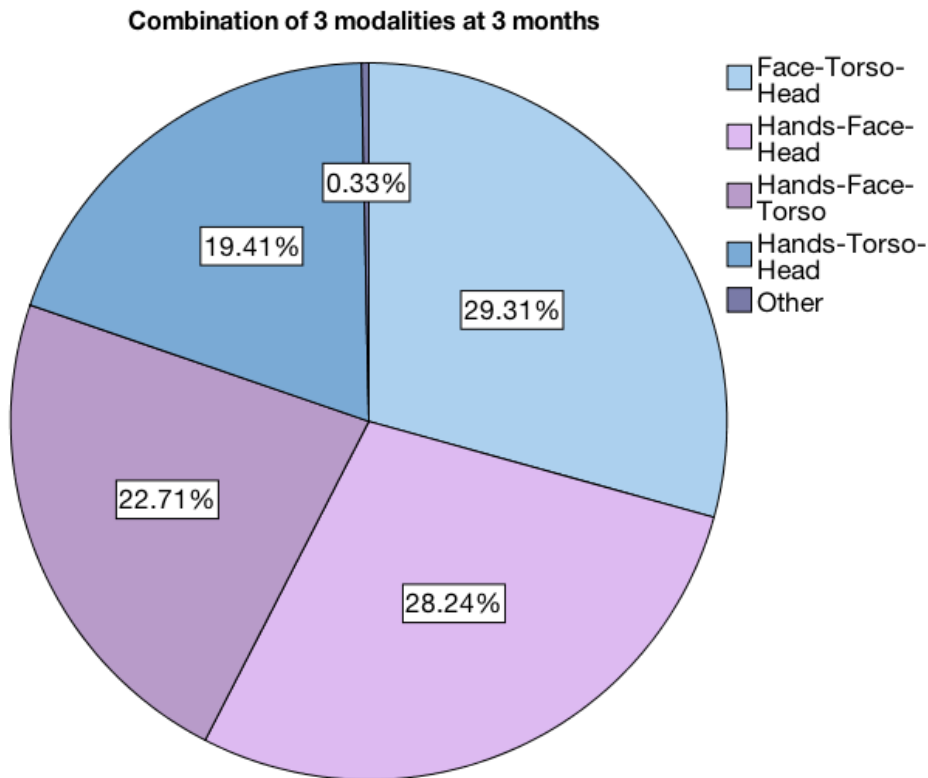
Figure 6.34.: Frequency of co-occurrence of three modalities at three months of age

Six-month-old infants

Here again, with the six-month-old infants there were differences in the extent to which three modalities co-occurred, accompanying the verbal behavior of the mother. Most predominantly, the HANDS-FACE-TORSO combination was used–which was the third most frequent for the three-month-olds–occurring for an average of 3.77 % (*SD* = 3.55) of spoken time. This was

followed by FACE-TORSO-HEAD (2.28%; *SD* = 2.62)–which was the most frequent combination for the three-month-olds , and HANDS-FACE-HEAD (2.02% ; *SD* = 2.41)–which was the second most frequent combination for the three-month-olds. Finally, HANDS-TORSO-HEAD remained in fourth place (1.57% (*SD* = 1.75 ; *MIN* = 0 %; *MAX* = 6.43%). In this case, the remaining modality pairs occurred less than 0.2% of the spoken time, but all possible modality combinations were encountered in the data. The frequency of occurrence of the co-occurring modality pairs is illustrated in Figure 6.35.
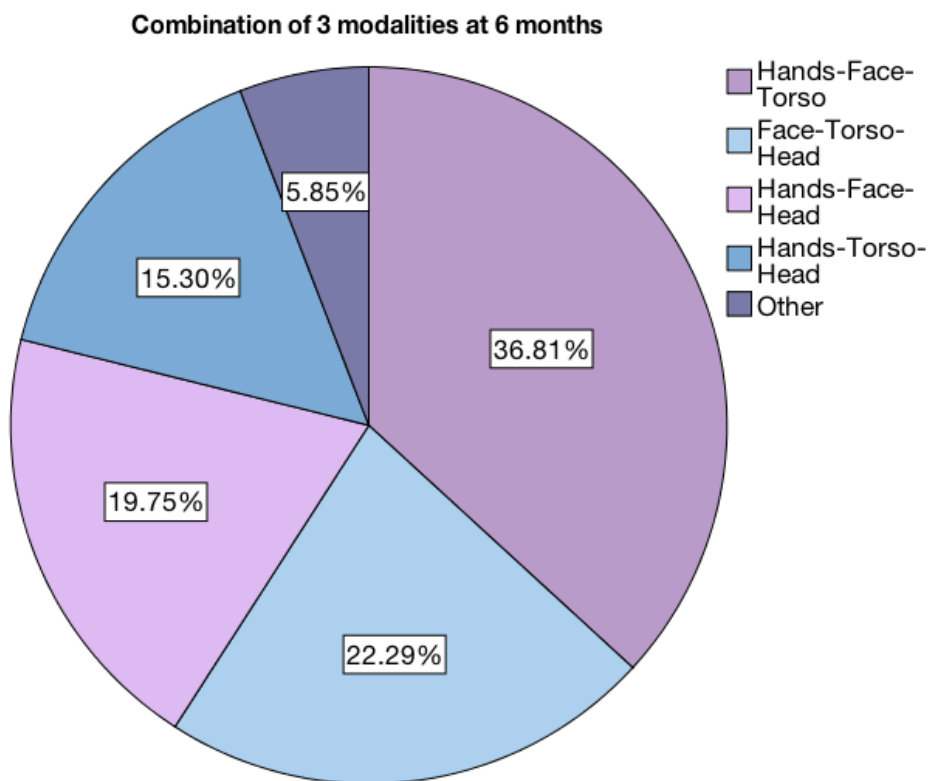
Figure 6.35.: Frequency of co-occurrence of three modalities at six months of age

Combination of four modalities and vocal behavior

For sets of four modalities there were 5 possible combinations. The descriptives for these combinations can be seen in Appendix Table B.18 and Table B.19 for the three-month-old and six-month-old infants, respectively.

Three-month-old infants

Four-modality combinations were also found, the most frequent being HANDS-FACE-TORSO-HEAD, although the duration of overlap was very short (average of 0.26 % of spoken time, *SD* = 0.38). The other combinations were not encountered in the data.

Six-month-old infants

For the six-month-olds, there were overlapping intervals in the HANDS-FACE-TORSO-HEAD combination (0.79%; *SD* = 0.92). Overlapping intervals were found for all other possible modality combinations. Nevertheless, the durations of overlap comprised less than 0.05% of the spoken time.

Combination of all modalities and vocal behavior

Finally, there was only one event found in the interaction of a single mother with her six-month-old infant, in which all five modalities overlapped.

6.4.8. Comparison of modality co-occurrence in the two age groups

To test whether age influences the use of co-occurring modality pairs, a repeated measures ANOVA was calculated for the combination of two, three, and four modalities.

   For the co-occurrence of two modalities the ANOVA revealed a main effect for AGE, $F(1, 16)$ = 13.66, p = .002, η^2 = .461, and and interaction effect for AGE x OVERLAP, $F(1, 16)$ = 3.80, p = .037, η^2 = .810. Post-hoc pairwise comparisons revealed an overall increase of co-occurring

modalities for the six-month-olds. A paired sample *t*-test showed a significant increase for the pair HANDS-FACE: $t(16) = 4.62$; $p < .001$, the pair HANDS-TORSO: $t(16) = 4.86$; $p < 0.001$, for FACE-TORSO: $t(16) = 3.33$; $p = 0.004$, i.e the most frequent modality pairs for the six-month-olds. A significant increase was also observed for following pairs: FACE-LIPS: $t(16) = 2.33$; $p = 0.03$ and HEAD-LIPS: $t(16) = 2.54$; $p = 0.02$.

For the co-occurrence of three modalities, the ANOVA revealed a main effect for AGE, $F(1, 16) = 11.55$, $p = .004$, $\eta^2 = .42$. As in the case of the combination of two modalities, post-hoc pairwise comparisons revealed an overall increase of co-occurring modalities for the six-month-olds. More specifically, there was a significant increase for the combination HANDS-FACE-TORSO: $t(16) = 3.15$; $p < 0.01$, HANDS-TORSO-HEAD: $t(16) = 2.98$; $p < 0.01$ and TORSO-HEAD-LIPS: $t(16) = 2.30$; $p < 0.05$.

When comparing the overlap of four modalities for the two separate visits there was, again, an overall increase of co-occurring modalities for the six-month-olds. This increase is evident in the fact, that only one combination type was found in the data of the three-month-olds, whereas in the data of the six-month-olds all possible combinations of four modalities were represented. The paired sample *t*-test showed a significant increase for HANDS-FACE-TORSO-HEAD: $t(16) = 2.44$; $p = 0.03$.

As already reported above, there was only one case of an interval in which all modalities and verbal behavior overlapped. This was in the data of the six-month-olds.

The results from this analysis, underline an important fact: Infant stimulation in natural conditions not only involves bimodal vocal-motor synchrony, but this redundancy in stimulation is further reinforced by synchrony across more than two modalities, rendering it genuinely multimodal. Also, it was shown that the use of overlapping modalities increases with age.

6.4.9. Summary of findings of the quantitative analysis

In the quantitative analysis the first question asked was to what proportion of the interaction with the infants do mothers employ multimodal practices in the form of coordinated vocal behavior

with body movement. The quantitative results revealed that irrespective of modality, multimodal stimulation occurred for an average of 29.1% of the entire interaction time when the infants were three months old, and 27.06% of the interaction time when they were six months old, with some mothers reaching 43.87% and 44.25 % of the interaction time, respectively. The most prominent modality used in synchrony with vocal behavior was that of the hands. This was observed in both visits. When comparing the two visits at the families' homes descriptively, a change in the frequency of use of the modalities is observed. While for the three-month-olds the hands are followed by the head, the face, the upper body, and finally the lips, for the six-month-olds the hands are followed by the face and the upper body, and the use of head follows these alongside with the lips. A test of the influence of age on the occurrence of multimodal practices revealed an interaction between age and modality. More specifically, the overall use of synchrony did not change with age. However, which modalities convey the synchrony proved to be age-dependent. More specifically, in comparison to interacting with three-month-olds, when interacting with six-month-olds, mothers (a) significantly increased the use of the facial modality, while significantly reducing the use of the head modality, and (b) significantly increased movements of their upper body (looming in and out).

Next, a comparison was made between the duration of synchrony events at the two observation times. This revealed a main effect for age, suggesting that as the infant becomes older the synchrony events used by the mother became shorter. Here, also an interaction was found between age and duration of synchrony events.

Furthermore, a question was raised about how much of the entire speech was of a multimodal nature. This would give a more precise measure of the extent to which verbal behavior is reinforced by body movement. In the interactions with the three-month-olds, mothers vocalized for an average of 39.38 % of the interaction time. Out of this vocal behavior, an average of 73.34% was performed multimodally, with maximum values reaching up to 95 %. In the interactions with the six-month-olds, mothers vocalized for an average of 36.14% of the interaction time, and, on average, 69.73% of all vocal stimulation was performed multimodally. Here again,

maximum values reached 94%. The results speak for a very strong multimodal reinforcement of the vocal signal.

The next analysis step investigated the extent to which infant stimulation involved not only bimodal vocal-motor synchrony, but a redundant stimulation through synchrony across more than two modalities. For this, the proportion of mothers' vocal behavior was calculated, during which more than one synchrony modalities co-occurred. Indeed, combinations of two, of three, and four modalities were found in the data. Amongst them, certain combinations were more prominent than others. For example, in the interactions with the three-month-old infants, mothers reinforced their verbal behavior most frequently with overlapping movement of the head and the activation of the facial modality (an average of 7.26 % of the spoken time was taken up by this combination with maximum values reaching 15.75 %). In the interactions with the six-month-old infants, the most prominent combination was that of torso movements overlapping with hand movements (an average of 9.89 % of spoken time with maximum values reaching 24.90 %). All possible combined pairs of modalities were encountered in the data at both observation times. Also, overlaps of three modalities were found in the data. In the interactions with the three-month-olds, the most frequently occurring combination was that of the head and face modalities co-occurring with torso movements. For the six-month-old infants, mothers used a combination of the facial modality together with the hands and torso modalities the most. In the interactions with the six-month-old infants, mothers used all the possible combinations of modalities. This was not the case for the interactions with the three-month-olds in which only five out of ten combinations were encountered. Even combinations of four modalities were encountered, although very scarcely. Here, the most prominent combination was the one excluding the lips modality, i.e., hands-face-torso-head. When testing the influence of age on the occurrence of multiple overlapping modalities, a main effect for age was found for the combination of two and three modalities. As the infants became older the mothers provided their verbal behavior in a more overlapping manner.

6.4.10. Vocal types classes

Since the analyses focus on the synchrony between action and language, a further aim was to provide a detailed description of the vocal practices that mothers used while interacting with their infants. Besides identifying vocal–body movement synchrony, it seemed interesting to reveal the verbal content of such multimodal maternal utterances. Research on maternal vocal input has stressed its significance for the development of affect and attention (Trevarthen, 1979; Fogel, 1988; Legerstee, 2005) and has identified its possible functions for gaining and holding the infant's attention, establishing an affective bond between the infant and the caregiver, and allowing the earliest communication between them to take place (Sachs, 1977).

M. Papoušek (1994) points out that most existing work on the analysis of the baby talk register (i.e., speech directed toward young children) has focused mainly on linguistic methods and included mostly the vocabulary, the syntactic structure, and the phonology of infant-directed speech. She stresses that this practice ignores the special features of early motherese, such as prosody and responsiveness to infant vocalizations. Following the perspective of M. Papoušek (1994), all maternal vocalizations were treated as communicative vocalizations. Vocal type categories were developed to account for them (see Table 6.3). Thus, elements – which have elsewhere been characterized as "nonsense" (Toda, Fogel, & Kawai, 1990, p. 287), "paralinguistic information" (Henning, Striano, & Lieven, 2005, p. 524), or "funny noises" (Kaye, 1980, p. 493) – were treated as language, because they were coordinated meaningfully with action. In the following, the term 'language' is used to account for all vocal activities described in Table 6.3, and 'speech' as only a subcategory of this vocal behavior.

The present section includes:

 a) qualitative examples for all vocal types

 b) a quantitative description of the frequency of occurrence of vocal types for both points of observation

 c) a comparison of the occurrence of vocal types in the two age groups

d) an analysis of potential relationships between types of body movement and vocal type classes

Coding was carried out as a complementary step on existing multimodal segments. After coding the data for the body movement classes, all annotations were viewed and allocated to one of the five vocal type classes listed in Table 6.3. In the case of overlapping vocal types, for example when, within a single annotation interval the mother used speech and exaggerated breathing, both types were annotated.

6.4.11. Results of vocal types

In the following, the results are presented for the classification of annotations in *'vocal types'*. It has been suggested that early mother-infant interactions are "content-free" (Fogel, 1988, p. 393), and that they serve as a setting "for infant development of affect and attention" (Fogel, 1988, p. 393). This has been pointed out by other researchers, too.

Kaye (1980, p. 491) proposes that:

> "Rather than trying to make herself understood, the mother is trying to make herself interesting to her infant."

Sachs (1977) believes that if infants have a perceptual sensitivity to the various characteristics found in early input, then this input may function in:

- gaining and holding the infants attention

- establishing the affective bond between the infant and the caregiver

- allowing the earliest communication between them to take place

As part of the analysis of vocal-motor synchrony in early mother-infant interactions, it was deemed important to identify and describe the different types of vocal stimulation, which the

Table 6.3.: Vocal Types and their Description

Vocal Type	Description
Speech	Includes mother's utterances as well as onomatopoeic expressions such as (German) "Kuck Kuck" in a peek-a-boo game;
	other interjections and transliterated sounds such as "poing" or "shwoop", usually accompanying relevant actions;
	but also fillers and sounds such as "aha", "huh", and "aiaiai"
Emotionalization	Includes laughing, intentionally exaggerated breathing, or sighing usually accompanied by appropriate facial expressions or the expression of emotional states such as sadness or surprise.
Imitation	Includes the cases in which mothers imitate their infant's vocalization. This category was coded in cases in which there was an infant vocalization directly beforehand. This was considered a reliable measure for judging whether the mother intentionally imitated the infant or merely vocalized (in which case it would be coded as speech and not imitation).
	In some special cases, the mothers also used a certain vocalization, which had occurred previously to instigate a new interactional episode. This would also be coded as imitation, provided the vocalizing sound had been uttered before and was part of a routine between the specific mother–infant pair.
Sound	Was used for click or flick sounds, sometimes produced by the mother to get the infant's attention
Song	Baby songs or other improvised song-like speech sequences

mother chose in order to fulfill the above mentioned aims. Interactions were observed to have a reciprocal dialogical character, resembling that of turn-taking in adults, with alternating turns, or overlaps. Nevertheless the content of the utterances was of a different kind. Sachs (1977, p. 61) suggests that:

> "[...]This may allow the child to enter into a 'dialogue' very early with his caregiver, long before he has words or sentences. By looking at these mutual 'babblings', we may be able to trace communicative interactions back long before the onset of speech and see the roots of language in the 'prelinguistic' period."

Mothers generally addressed their infants as if they were equal conversational partners.

Example 30

Utterances such as:

```
"Gib mir das Handtuch"
```
"Give me the towel"

implicate that the infant is actually capable of understanding the mothers request and in position of intentionally handing over an object.

Example 31

Also there were many interrogatives such as,

```
"Hase oder das?"
```
"Do you want the bunny or this?"

which presuppose the infant capacity to answer. Yet mothers, of course, know what their infants can or cannot do, and are just pretending that they can. Toda et al. (1990, p. 279) hold that this can be explained by the fact that,

"Maternal speech to prelinguistic infants may serve dual purposes, as input for language acquisition and as socialization for culturally appropriate communication."

.

The topics of conversation which were observed in the present data were also as identified by Toda et al. (1990). They primarily concerned the infant's actions and his or her feelings. Sometimes the mothers referred to the external environment, i.e., the room in which the interaction took place, the furniture, the objects used for diaper changing, the view outside the window, and also the mother's behavior.

6.4.12. Qualitative description of vocal type classes

Speech

The observed maternal utterances had the characteristics of Infant-Directed Speech (Ferguson, 2004)[2] [3], (Kaye, 1980)[4].

Kaye (1980) has proposed a distinction between two levels of Baby Talk, one which he called BT1 and which was observed during the first 12 months (addressing pre-verbal infants), and

---

[2]Infant directed speech is modified speech addressed to infants. The modifications are prosodic (e.g. higher pitch), grammatical (e.g. shorter sentences), lexical (e.g. special baby words), phonological (e.g. reduplication), and discoursal (e.g. greater proportion of questions).

[3]Ferguson refers to it as "Baby Talk Register" in the sense that it is a set of characteristics, distinguishing language by social context and not by linguistic community.

[4](Kaye, 1980, p. 489) describes the characteristics of such language as "fall[ing] into five general classes:

    a) prosodic features [...]

    b) lexical features [...]

    c) complexity features - shorter utterances [...]

    d) redundancy features - more immediate repetition [...]

    e) content features - restriction to topics in the child's world [...]"

another which he called BT2, which began after the 1st year. The speech observed in the specific data resembled the BT1 proposed by Kaye (1980). It consisted to a great extent of very short, repetitive utterances. Also lexical baby talk did not occur very often. It was more the temporal (duration of phrases) than the lexical (baby words) features and the highly repetitive character of speech which made it *'babyish'*. This aspect of maternal speech was illustrated in Transcript 17 for instance. There is was shown how this repetitive verbal behavior was accompanied with torso movements.

Example 32

In this example the mother uses a repetitive format, as she repeats the same phrase in lines 1, 2 and 4.

VP09 4tht visit (01:23–01:40)


```
1 M: hinten dran und einschlagen ((inbreathe)) dann machen=
   fix on the back and fold it ((inbreathe)) then we=
   wir den Body zu (0.7 s)
   we close the onesie (0.7 s)
2 M: dann machen wir den Body zu (0.7 s)
   then we close the onesie (0.7 s)
3 I: ((laughing))
4 M : [machen wir den Body zu? (0.9 s)
   do we close the onesie? (0.9 s)
```


This repetitive theme is interrupted by the mother commenting on the infant's laugh (line 6). She then resumes the use of a repetitive format using another theme (lines 8, 10 and 12).

```
5 I : [((laughing))
6 M: ja:: [findest du das lustig? (1.1 s)
   ye::s do you find this funny(1.1 s)
```

```
7 I: [((laughing))
8 M : nasenstüber (0.7 s)
  tap on the nose (0.7 s)
9 I: ((laughing))
10 M: nasen[stüber (0.7 s)
  tap on the nose (0.7 s)
11 I : [((laughing))
12 M : nasenstüber (0.7 s) hä?
  tap on the nose (0.7 s) huh?
```

Furthermore, the speech observed was "more like speech to adults than BT2" (Kaye, 1980, p. 501). It might actually be that "[p]arents do not make up babyish forms until the baby is old enough to understand them" (Kaye, 1980, p. 501) or to produce them. Toda et al. (1990) also suggest, that maternal speech to young infants is more complicated than that to older infants. They observed that "when mothers talk to prelinguistic infants, their speech is more conversational than it is to infants who are using single words" (Toda et al., 1990, p. 281). This was the case in Transcript 2.

Example 33

VP10 1st visit (00:00-00:16)

```
1 M: ((imitation)) (0.6 s) ziehen wir jetzt mal den=
  =Schlafsack aus? (1.1 s)
  ((imitation)) (0.6 s) shall we take the sleeping bag=
  =off now? (1.1 s)
2 M: he? sollen wir den Schlafsack ausziehen? (1.5 s)
  hm? should we take off the sleeping bag? (1.5 s)
3 M: ((inbreathe))(2.2 s)
```

```
4 M: so (- - - ) guck mal die Oma hat eben vergessen den Knopf=
  = zu zu machen (.) he? (2.3 s)
```
so (—) look granny forgot to close the button before=
=(.) huh? (2.3 s)

Here the mother is speaking in long sentences, asking for the infant's acknowledgement by using *"huh"*.

A further type of speech which occurred very often in the present data and complies with the BT1 assumption were phatics, i.e., one word greetings, fillers or sounds, e.g. *"yes"*, *"hello"*, and *"oh"*. They were usually produced in response to the infant's attention or expressive action. This was illustrated in Transcript 9.

Example 34

Here the mother reacts to the reestablishment of mutual attention with the infant, by using a greeting (line 2).

VP09 4th visit; (03:54–03:57)

```
1 M: räumen wir gleich weg
```
we shall put this away in a moment
```
2 M: (0.5 s) hallo du
```
(0.5 s) hey you
```
3 I: ((sigh))
```

Kaye (1980) found that 21% of the utterances in mothers' interactions with one-to-three-month-old infants were one-word greetings or what he called "acknowledgements of real or imagined communications from the baby" (Kaye, 1980, p. 500). He observed that such feed-back expressions as *"yeah"* or *"sure"* are more common in adult-to-adult speech, since they are used when another person has the floor. Considering the fact that these phatic utterances

were found to be rarer in the speech to 2-year olds (who have in fact started to speak and are capable, in some extent, of holding the floor), he hypothesized that they are common in speech to preverbal infants, because the mothers pretend that the infants have the floor. Again, this type of utterance occurred very often in the present data. Kaye (1980, p. 490) concludes that, in accordance to the suggestions posed by C. E. Snow (1977),

> "the features of baby talk are due to the mother's attempts to maintain the resemblance of a conversation despite the conversational deficiencies of their partners."

Example 35

In this example the mother uses a feedback signal to acknowledge the infant vocalization. The sequence begins with the mother commenting on the infant having put her toes in her mouth (lines 1-3). In line 4, the infant vocalizes. The mother directly responds to this vocalization with *"u-huh"*.

VP17 4th visit; (03:08–03:28)

```
1 I: [((sigh))
2 M: [a:::h die Zehe die ist auch lecker, hm?
   o:::hh the toe, that is also yummy, huh?
3 M: lecker, lecker, lecker, lecker (2.2 s) mmm (4 s)
   yummy, yummy, yummy, yummy (2.2 s) mmm (4 s)
4 I: ((vocalization))
5 M: aha?
   uh-uh?
6: I: ((vocalization))
7 M: ((imitation)) (- - -) ganz toll
   ((imitation)) (- - -) super-duper
```

A further prevailing subgroup of the speech category was that of onomatopoeic expressions and transliterated sounds, e.g. *"pitsch"* or *"plop"*. This category has mostly been excluded from the speech category or labeled as "nonsense" (e.g., Toda et al., 1990). However, in the present data it emerged as a type of speech accompanying actions. The setting of the present study involved a specific activity and mothers were observed to produce such onomatopoeic expressions in synchrony with their actions, as if making sound effects for the actions carried out. It seemed as though mothers were trying to preserve the multimodal redundancy between speech and body movement, and when the body was involved in a manual task, they provided the redundancy by verbalizing their actions in an onomatopoeic manner. This was illustrated in Transcripts 13, 14 and Example 20, in which the mothers accompanied their actions with the interjections *"yippy"*, *"plup"*, and *"hui"*, respectively.

Gogate et al. (2001) hypothesize that the detection of arbitrary relations between co-occurring events that are predictable in nature may be easier than the detection of the arbitrary relations between less predictably co-occurring events. They believe that pairings resulting from natural relations, such as an action and the sound an action makes, may be more predictable and, therefore, more facilitative for infants during the detection of word-referent relations than pairings of arbitrary relations, such as that of an action and a verb. For older infants Imai, Kita, Nagumo, and Okada (2008) have shown that sound symbolism actually facilitates verb learning. This raises the question whether mothers' use of onomatopoeic and verbalized sounds may also be serving the same purpose. Even if the infants are too young to actually detect such arbitrary relations, it could be that mothers are intuitively reciprocating the infants' ability to do so.

Imitation

Mothers imitated infant vocalizations often in the data. The imitation was employed in different ways. Often the mother would imitate a vocalization with some exaggeration. This included adding multiple modalities while mirroring the vocalization of the infant. Also, as illustrated in Transcript 10 the exaggeration may involve imitating the vocalization multiple times. Further-

209

more, imitations of infant vocalizations were conversationally contextualized. For example, the mother would repeat the infant vocalization and then proceed with her utterance, as if she was acknowledging the vocalization as a contribution to the conversation and elaborating on it. This was illustrated in Transcript 2 (see example below).

Example 36

VP10 1st visit (00:00-00:05)

```
1 I: ((vocalization))
2 M: ((imitation)) (0.6 s) ziehen wir jetzt mal den=
   = Schlafsack aus?
```

   ((imitation)) (0.6 s) shall we take the sleeping bag off now?

Furthermore, mothers would imitate infant vocalization altering the intonation pattern making it sound like a question. This strategy also made the infant vocalization seem embedded in the ongoing sequential organization of the conversation. A further instantiation of this use of imitation involved the alternation of multiple vocal turns between mother and infant in which the mother always repeated the infant's last vocalization. Finally, mothers imitated not only infant vocalizations but also non-speech sounds such as sneezes or whimpers. These non-speech sounds were also sometimes verbalized.

Example 37

For example, when the infant sneezed, the mother would imitate by using the onomatopoeic expression *'hatschi'*.

M. Papoušek and Papoušek (1989) stress the supportive role of mothers' imitation of the baby[5] to their learning of vocal imitation. By imitating infant vocalizations, mothers "provide models which may most easily elicit matching responses" (M. Papoušek & Papoušek, 1989, p. 149).

---

[5]M. Papoušek and Papoušek (1989, p. 149): "It is apparent [...] that some of these [vocal] adjustments occur as immediate, exaggerated matches of features used in infant vocalizations. [...] [M]aternal adjustments function as age-appropriate models with respect to basic articulatory skills that the infant needs to practice [...]."

They believe that,

> "The majority of mothers seem to act as if the baby had the capacity to imitate, by providing a modeling-echoing frame around infant vocalizations. The maternal frame offers the infant abundant opportunities to experience reciprocal vocal matches long before the infant may be able to imitate. The matches allow intermodal integration of sounds heard and seen from the mother with those produced by the infant, thus paving the way for learning how to imitate."

In the data included in the present analysis, such maternal imitation of infant vocalizations and behaviors was coded when it was found to be multimodal.

Sound

Sounds such as whistles or clicks were considered as significant attention-getting signals and were observed to be synchronized with movement, so they were included in the analysis but formed a category of their own. The use of sounds was illustrated in Transcript 1.

Example 38

In this example, the mother starts by commenting on the fact that it will be worth changing the diaper. She begins a next utterance but discontinues as the infant vocalizes. As a response to the infant vocalization the mother produces a sound.

VP17 1st visit (00:33-00:38)

```
3 M: ja [wa
  yes [wha
4 I: [((vocalization))]
```

(a)   (b)   (c)
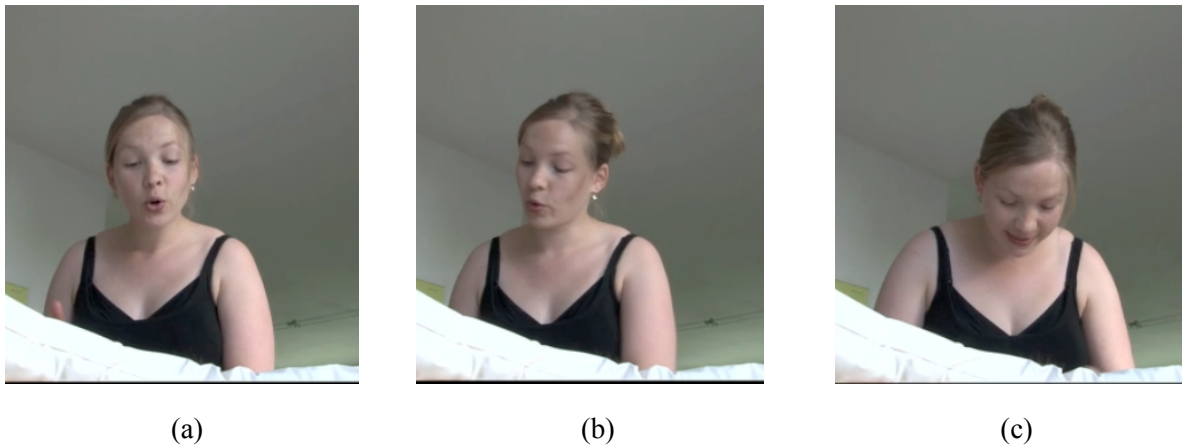
Figure 6.36.: Mother reinforcing sound vocal type with the facial modality

```
5 M: *ᵃ ((sound)) *ᵇ (.)((sound))(.)guck mal *ᶜwas jetzt kommt
     ((sound)) (.) ((sound)) (.) look what is coming now
```

In this example, the mother uses the sound vocal practice in synchrony with the facial modality (see Fig 6.36). Sounds were also illustrated in Example 29, in which the mother produced a flick sound, while at the same time raising her eyebrows to further reinforce her vocal practice.

Emotionalization

This category includes a variety of non-speech sounds produced by the mother, including exaggerated breathing, panting, or sighing. During coding of vocal-motor synchrony, it was observed that mothers frequently used an audible, exaggerated inhalation when interacting with their infant. It is further speculated that it could also serve as a kind of filler between phrase pauses, as illustrated in the following example.

Example 39

Here, the mother utters a phrase and fills the pause between the first and second phrase with a loud inbreathe accompanied with a head movement.

VP17 1st visit (00:35-00:40)

```
5 M: ((sound)) (.)((sound))(.)guck mal was jetzt kommt
```
((sound)) (.) ((sound)) (.) look what is coming now

(a)   (b)
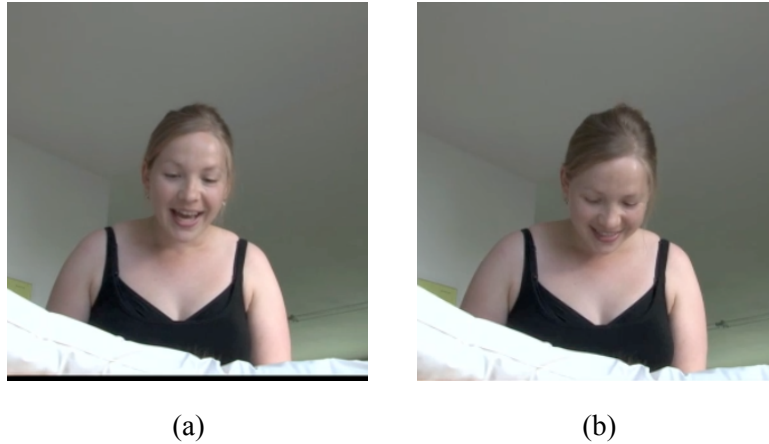
Figure 6.37.: Mother using emotionalization with upward head movement

```
6 M:((inbreathe)) *a was kommt jetzt? *b (1.4 s)
```
((inbreathe)) what is coming now? (1.4 s)

The frequency of occurrence of such emotionalization signals led to the formation of a separate category for this type of sound. It very often co-occurred with one of the synchrony modalities. Zukow-Goldring (1997) calls this behavior an "inbreathe" (Zukow-Goldring, 1997, p. 229) and has shown it accompanies preparatory actions for mother's stimulation. This was illustrated in Transcript 10 in which the mother is massaging the infant's body.

Example 40

Here the mother produces a loud inbreathe before starting to massage the infant's shoulder.

VP10 4th visit (02:50-02:54)

```
1 M: ((inbreathe)) über die Schulter
     ((inbreathe)) over the shoulder
2 M: (0.9 s) und die andere Schulter
     (0.9 s) and the other shoulder
```

It is as though mothers use this as an introduction, like saying *"Listen! I am going to do something now."*

This was also shown in Transcript 13 in which the mother sighs before lifting the infant off the changing table.

Song

Singing is very frequently employed in child care. In a diary study, Trehub and Trainor (1998) found that singing occurred not only in play situations (36%), but also in the course of routine care such as feeding (19%), diaper changing (6%), and bathing (6%), as well as during car travel (10%). Friedlander (1968) found that infants chose to listen to nursery songs three to four times as often as other kinds of verbal recordings. Furthermore, Trehub and Trainor (1998) report that play songs "were overwhelmingly the songs of choice (64%)" (Trehub & Trainor, 1998, p. 55). In these play songs body movement plays an important part. Suliteanu (1979, p. 205) proposes that they are designed

> "to amuse the child when he is awake by lifting him up in the arms, playing with
> his fingers and palms, tickling him, moving his hands and feet, teaching basic body
> movements"

Trehub and Trainor (1998) suggest that this type of songs alter their form and function over time. They believe that in early months of a child's life, "singers of play songs function, in part, as solo entertainers, optimizing infant arousal" (Trehub & Trainor, 1998, p. 55). This type of

song was identified in the present data. Interestingly enough, these singing interactions had a multimodal character, involving movement of the infant's limbs by the mother or accentuated movement of her head. Longhi (2009) also reports the existence of what she calls "songese", and also stresses the multimodal character of mothers' singing interactions with infants. Longhi (2009, p. 208) proposes that,

> "[...] the integration of the mothers' multimodal sensory information ensures a temporally coherent segmentation of the musical event while singing and moving their own as well as the infants' bodies."

The author observes that through their multimodal sensory information, the mothers emphasize the hierarchical structure of the song. This assists infants in the processing and segmenting of the musical event into smaller units and at different levels. Longhi (2009, p. 210) suggests "maternal speaking and [...] singing to the young infants are crucial in the process of learning language".

6.4.13. Quantitative analysis of vocal type classes

With the aim of assessing mothers' vocal behavior during the whole interaction with their children, the frequency of occurrence of vocal types in all modalities was calculated. To explore whether there were any changes in maternal verbal strategies between the first and fourth visit, vocal types were coded for both visits and compared.

Three-month-old infants

The results for the three-month-olds are presented in Appendix Table B.20. Frequencies of occurrence for vocal types were calculated.

Expressed in proportions, for the three-month-olds, speech was the most frequently occurring category in the data (72.10 %; $SD = 13.93$). Emotionalization (12.61 % ; $SD = 7.83$) was used less often, followed by sounds (8.67 %; $SD = 6.75$), and the imitation of infant vocalization (5.01

%; *SD* = 3.95). Singing was observed rarely (1.62 %; *SD* = 4.38 ; *MIN* = 0; *MAX* = 17.26 %) in the present data, but was still present as a vocal type.

Six-month-old infants

The results for the six-month-olds are presented in Appendix Table B.21.

Here again, for the six-month-olds the order from most frequently used to less frequently used remains the same, but the percentages are different. Speech was the most frequently occurring category in the data (61.79 %; *SD* = 20.58). Emotionalization (12.38 % ; *SD* = 6.30) was used less often, followed by sounds (12.36 %; *SD* = 8.29), and the imitation of infant vocalization (5.68 %; *SD* = 5.50). Singing was again observed rarely (1.90 %; *SD* = 6.29 ; *MIN* = 0; *MAX* = 26.15 %).

6.4.14. Vocal type comparison of age groups

The frequency of occurrence of vocal types is illustrated in Figure 6.38 and Figure 6.39 for the three- and six-month-old infants, respectively. There is a slight variation in the distribution of the vocal types. When the infants were six months old, the mothers' use of the speech vocal type decreased, in favor of the use of sounds. The remaining vocal types were used almost as frequently as in the interactions with the three month olds.

To explore whether there was a relationship between age and the vocal types used when providing multimodal behavior a cross-tabulation analysis was carried out. A chi-square test of independence was performed to test the null hypothesis of no association between age and vocal type.

The observed and expected frequencies for the distribution of vocal types at the two time points are presented in Appendix Table B.22. For SPEECH, the observed frequency at three months is bigger than the expected. The opposite is the case at six months, where speech occurred less than expected by chance. A somewhat bigger difference between observed and expected values is also found for the vocal type SOUND. In this case, at three months sound is used

Frequency of occurrence of vocal types at 3M
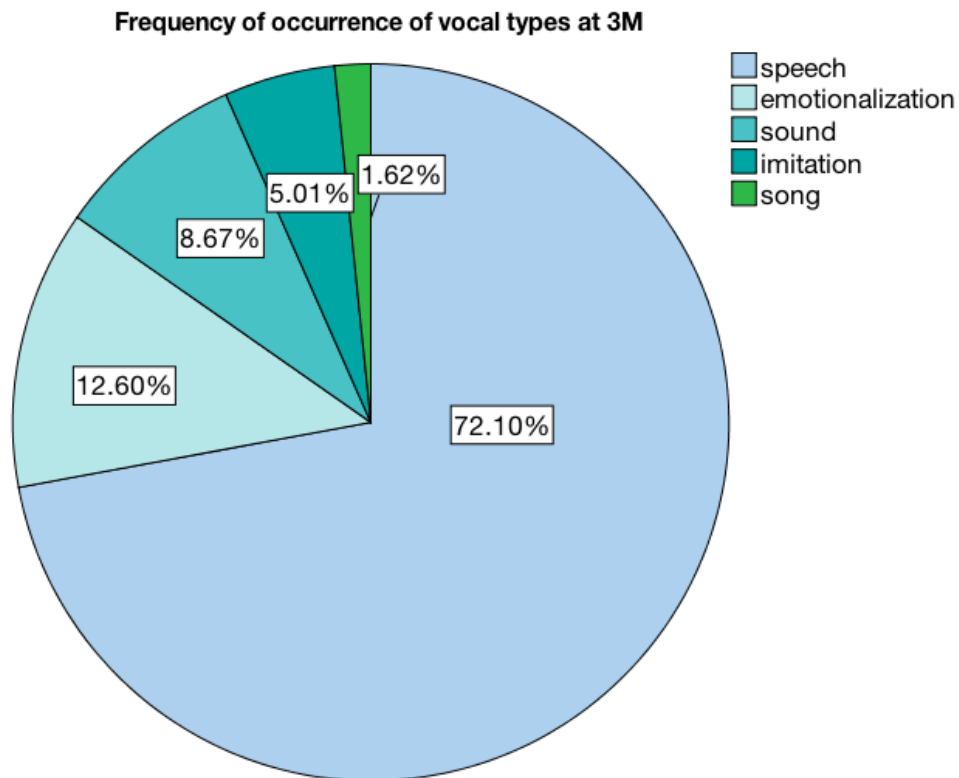
Figure 6.38.: Frequency of occurrence of vocal types at three months of age

less than expected by chance, and at six months sound is used more than expected by chance. Despite these observations an association between age and vocal type could not be statistically confirmed (χ^2 *(4, N = 412) = 1.48 ; df = 4 ; p = .83*).
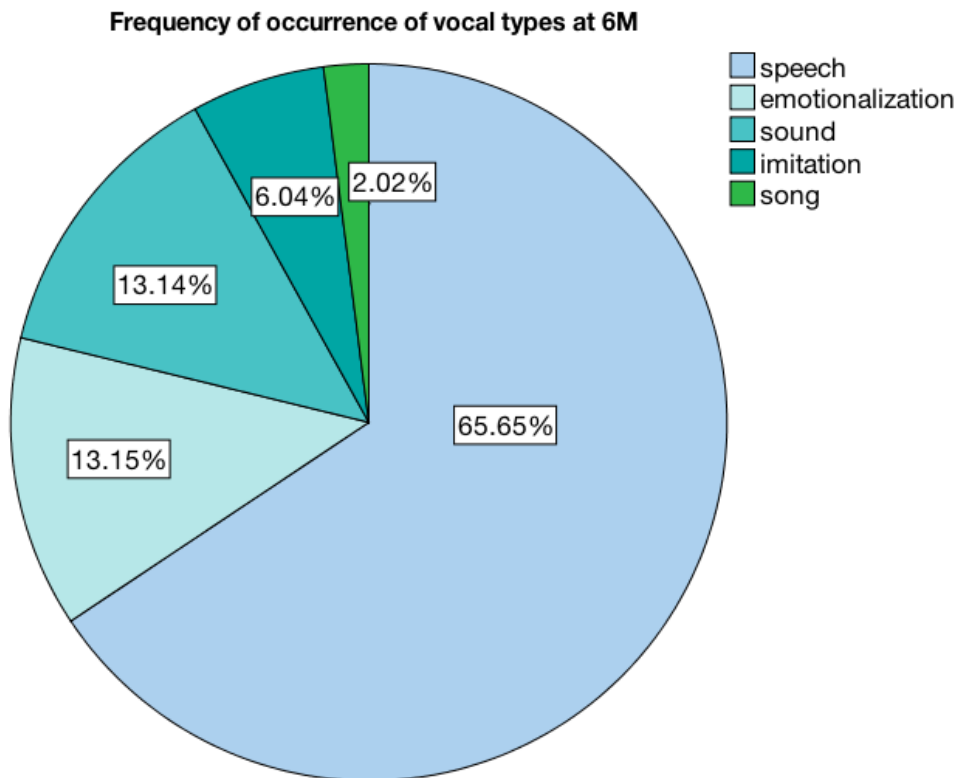
Figure 6.39.: Frequency of occurrence of vocal types at six months of age

6.4.15. Relationship between synchrony and vocal type classes

During coding of the data the impression arose that some body classes seemed to be systematically employed with certain vocal type classes. This gave rise to the question, whether specific types of body movement are more likely to occur with specific types of vocal behavior. To answer this question, a cross-tabulation analysis was carried out to investigate potential relationships between these two variables.

Three-month-old infants

The observed and expected frequencies for the combinations when the infants were three months old are presented in Appendix Table B.23. For FACE and EMOTIONALIZATION the observed frequency was higher than the expected, whereas, for example, in the case of FACE and SPEECH the observed frequency was lower than the expected. The same was found for FACE and IMITATION. Moving to the body class HANDS, mothers were observed to employ this modality in combination with the vocal type SPEECH more often than expected by chance. With regards to the body class HEAD, it was observed to occur with IMITATION more often than expected by chance. Finally, LIPS occurred almost exclusively with SOUND.

Six-month-old infants

The observed and expected frequencies for the combinations when the infants were six months old are presented in Appendix Table B.24. The potential relationships between variables remain the same. FACE and EMOTIONALIZATION and FACE and IMITATION show observed frequencies higher than the expected. Again, HANDS and SPEECH are observed very often, as well as HEAD and IMITATION. Finally, LIPS occurred almost exclusively with SOUND.

Age group comparison of synchrony-vocal type relationship

In the cross-tabulation analyses presented above, the same patterns emerge for both visits to the families' homes at three and six months. Furthermore, a chi-square test could not be calculated to test these observations, because of the zero frequencies observed for some combinations.

6.4.16. Summary of results on vocal types

The analysis of vocal types provided insights into the type of vocal strategies which mothers use when engaging in multimodal practices: The most frequently used was that of speech. This vocal type included known aspects of infant-directed speech such as short utterances, repetitions,

and many interrogatives, but also other practices such as phatics, interjections and onomatopoetic expressions. The second most prominent vocal type was that of emotionalization signals. These included panting, sighing, laughing and exaggerated inbreathes. Next most frequently employed vocal strategies were sounds, such as whistles or clicks, followed by imitations of infant vocalizations. Finally, mothers used singing that was coordinated with their body movement. A comparison between the two times of observation did not reveal any significant differences in the use of the various vocal types. Nevertheless, when looking at the data descriptively, in the interactions with the six-month-olds there was an increase in the use of the sound vocal type and a decrease of speech. A further analysis investigated whether specific body classes were more likely to occur with specific vocal types. This was particularly the case for face and emotionalization, face and imitation, hands and speech, and finally for lips and sound. This was found for both observation points.

6.5. General Summary

The present section revealed that verbal input to the infants in natural interactions simultaneously contains action information. Five classes of bodily movements that were synchronized with vocal activities were identified. Also five vocal types were identified that were used during these multimodal practices. Using these coding categories, various multimodal practices were identified in which action (in the form of bodily movements) and language (in the form of different vocal types) temporally overlapped each other. It was shown that language and action form "multimodal packages" and structure each other. It was shown that this structuring is very rich and variable, as various forms of co-occurrence provided a differential structuring, or packaging, of actions or language. Yet, the input is not just a behavioral modification activated when interacting with the infant. There is great variability in infant directed behavior, suggesting that mother and infant co-construct the input. The extent to which this is encountered in the present data will be investigated in the analysis presented in the following chapter.

7. Responsive Synchrony

7.1. Objective

In the previous chapters it was hypothesized that by synchronizing verbal behavior with body movement, caregivers recruit infants' attention and assist them in binding information from different modalities. In Chapter 6 it was shown that mothers actually make ample use of vocalizations which are provided in a tight relationship with action, making language perceivable and tangible to the infants.

The question raised in this section is how the infant – as an active interaction partner – regulates the behavior of the mother, and how the multimodal input is collaboratively constructed within everyday activities. This question was already addressed in Chapter 5. In the single case qualitative analysis presented there, it was shown that mothers modify the modalities they use and position themselves and the objects involved in the interaction according to the infants' direction of gaze. In this Chapter, this analysis is expanded to include microanalytical methods, thus generalizing from single cases to the entire data corpus with statistical analyses.

a) How does the infant - as an active interaction partner - regulate the input he or she receives?

b) Is infant gaze a factor influencing maternal input?

c) If yes, to what extent?

7.2. Method

7.2.1. Participants and Data

For this analysis the gaze coding for the infant were used as well as the synchrony coding from the first visit. The coding methods producing this data for analysis have already been described in Chapter 3 and Chapter 6. All seventeen mothers and their infants were considered for this analysis.

7.2.2. Coding

More specifically:

- The coded synchrony classes for the three-month-old infants were used.

- The gaze data of the infant for the three-month-old infants were used.

- Using these existing data, the overlap of each synchrony class with infant gaze was calculated, and thus new coded intervals were generated.

- Finally, the average duration of these synchrony-gaze intervals was calculated.

7.3. Analysis and Results

7.3.1. Multimodal Input and infant gaze

The initial observation was that mothers modify the modalities they use and position themselves and the objects involved in the interactions according to the infants' direction of gaze. This was already elaborated in Transcripts 8 and 9 in which the mother attempts to recruit the infant's attention using multimodal practices.

Example 41

In this excerpt, taken from Transcript 9, the mother follows the infants line of gaze and lowers her upper body, positioning her head within the infant's gaze (image a). Once she has the infant's attention (image b), she starts making forward and backward torso movements in synchrony with her speech (images b and c).

VP09 4th visit; (04:07–04:14)

(a)   (b)   (c)
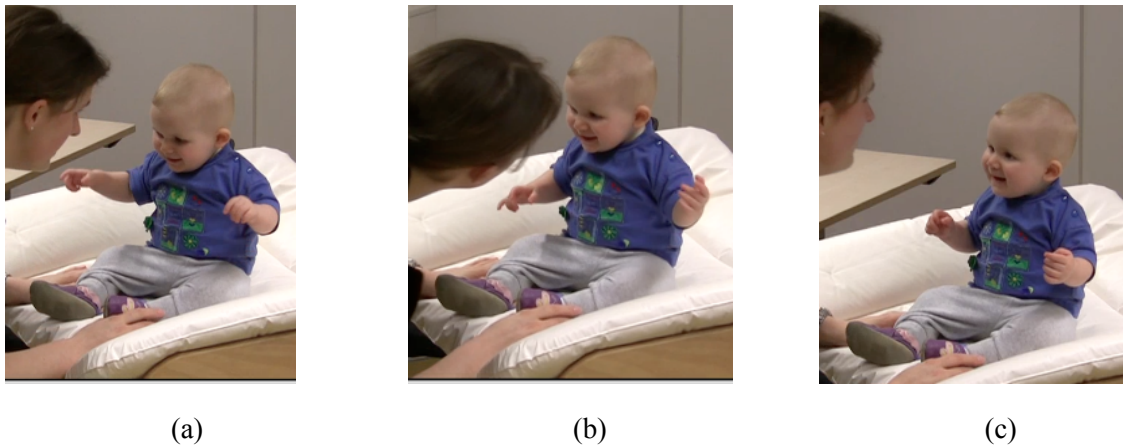
Figure 7.1.: Mother using synchrony according to the infant's attention.

To investigate the hypothesis that maternal multimodal practices are tailored to the infants' attention, the overlap of the synchrony provided by the mother with infant attention was calculated. More specifically, for each modality the duration was calculated during which each modality overlaps with the infant's gaze at the mother. Then, the duration of the overlap of each modality with the infant's gaze away was calculated. From these durations, the proportion of overlap of synchrony with the different infant gaze locations could be calculated (see Figure 7.2)

From this analysis it became evident that mothers synchronized their verbal behavior with their body movement more when the infant was looking at them than when the infant was gazing away (see Figure 7.3).

To test whether the multimodal maternal input is coupled to the behavior of the infant, a
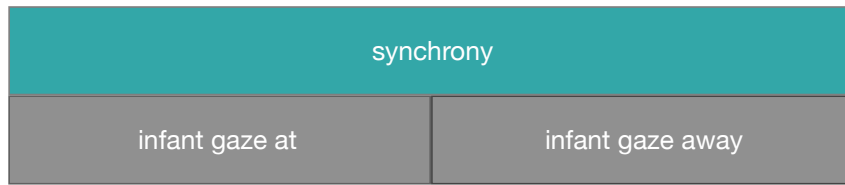
Figure 7.2.: Schematic illustration of the overlap of synchrony with infant gaze
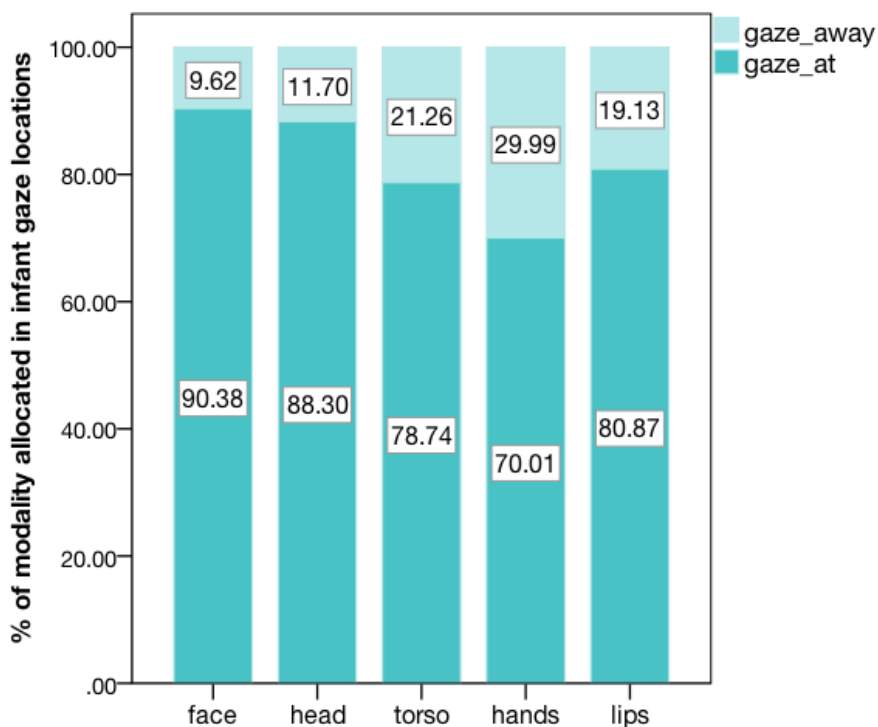
Figure 7.3.: Proportion of overlap of infant gaze locations with all synchrony modalities

repeated measures ANOVA was used. There was a main effect for Gaze $F(1, 16) = 84.14$; $p < .001$; $\eta^2 = .84$. This means that mothers synchronized speech and body movements more when the infant was looking at them than when s/he was looking away. Also, there was an interaction

effect for synchrony modalities X Gaze $F(4, 13) = 7.25$; $p = .003$; $\eta^2 = .69$. This suggests that the infants' direction of gaze affects all the modalities of synchrony provided by the mother. Post-hoc t-tests showed significant differences for all modalities. The results are presented in Appendix Table C.1.

7.3.2. Relationship between synchrony and gaze

The above results suggest that mothers track the attention of the infant when employing their multimodal practices. The input, thus, is interactively triggered and coupled to the infants' gaze. The next question to be investigated was whether the duration of the mothers' synchrony events matched the time the infants spent looking at them. This would reveal the extent to which mothers and infants co-construct the multimodal input.

Thus, the first results allowed the formulation of a new hypothesis: If the input is constructed for the infant's gaze, then there should be a relationship between the duration of synchrony events and the duration of infant gaze events. To test this, the average duration of intervals in which the infant's gaze overlapped with the synchrony classes was calculated and then this average duration of overlaps was compared with the average duration of events in the synchrony classes.

The data presented in Appendix Table C.2 show a strong correlation between the mean duration of gaze events at the mother and the mothers' synchrony events.

More specifically, for the HEAD and LIPS modalities there was very strong correlation between the duration of the mothers' synchrony events and the duration of the synchrony events provided when the infant was gazing at the mother. Furthermore, this finding was supported by a weak correlation between the mothers' synchrony events and the events provided when the infant was not looking at the mother. This is illustrated in Figure 7.4 and Figure 7.5, respectively.

Similar was the case of FACE (see Figure 7.6), in which there was no significant correlation (although marginal) for the overlap with gaze away. Thus, in the synchrony modalities which
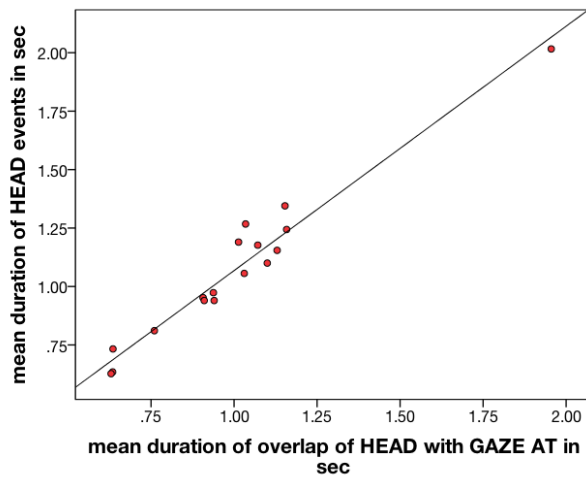
(a) Head vs. Head-infant gaze at
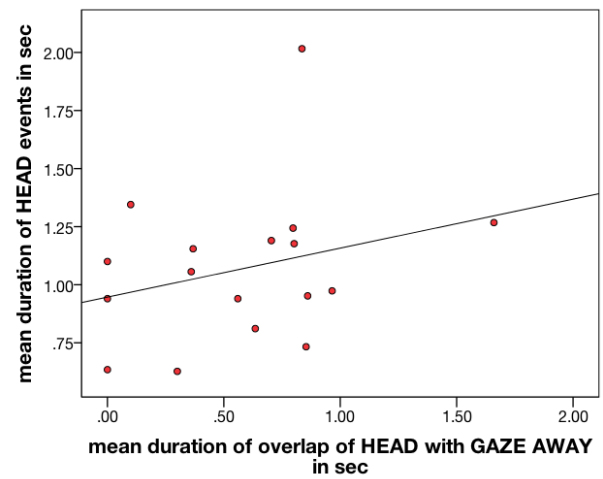
(b) Head vs. Head-infant gaze away

Figure 7.4.: Correlation between mean duration of events of infant gaze overlapping with the head modality and mean duration of events in the head modality
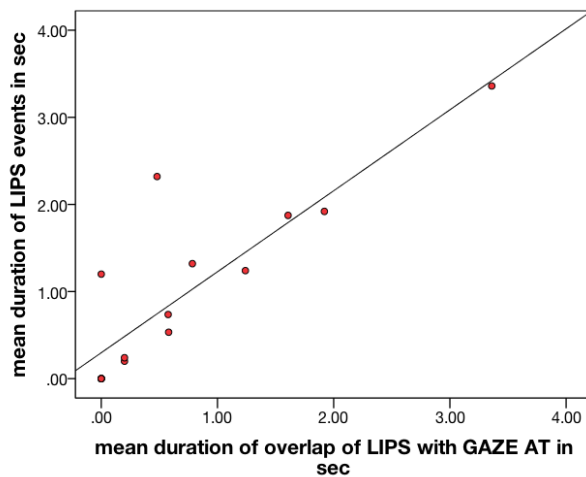
(a) Lips vs. Lips-infant gaze at
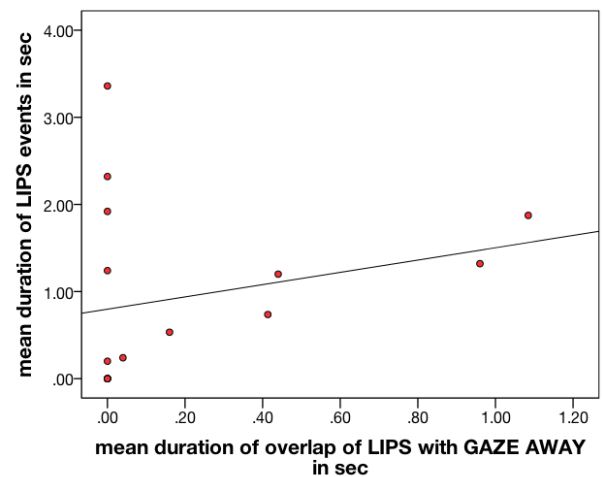
(b) Lips vs. Lips-infant gaze away

Figure 7.5.: Correlation between mean duration of events of infant gaze overlapping with the lips modality and mean duration of events in the lips modality

are more proximal to the eyes, the duration of the synchrony events seem to match the duration of the infant's gaze at the mother.
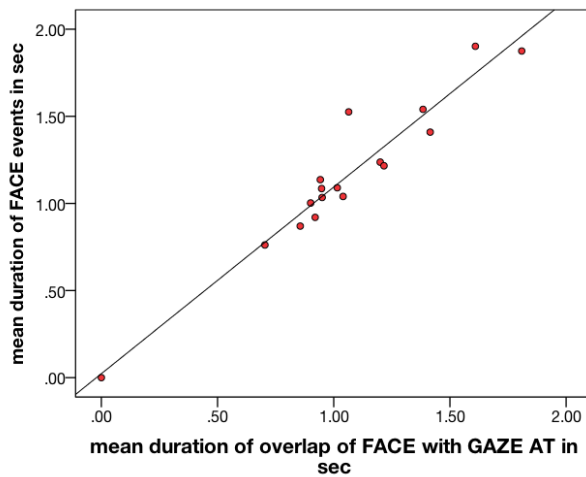
(a) Face vs. Face-infant gaze at
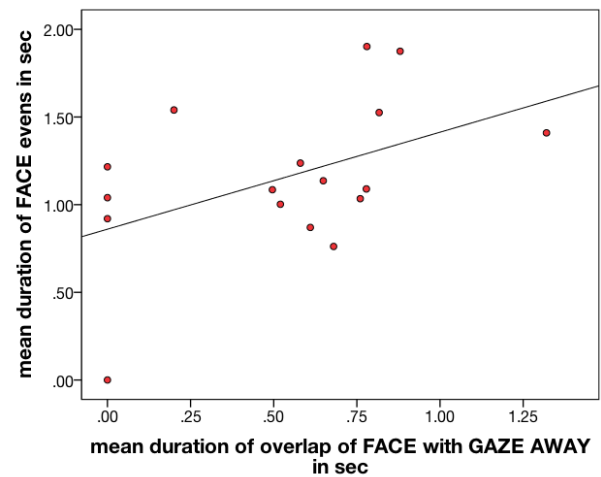
(b) Face vs. Face-infant gaze away

Figure 7.6.: Correlation between mean duration of events of infant gaze overlapping with the face modality and mean duration of events in the face modality

For the HANDS and TORSO modalities there was a significant correlation between the duration of the mothers' synchrony events and the duration of their overlap with infant gaze. Nevertheless, in this case there was also a significant correlation, although not as strong, between the duration of synchrony events and their overlap with the infant's gaze away from the mother's face (see Figure 7.8 and Figure 7.7, respectively.
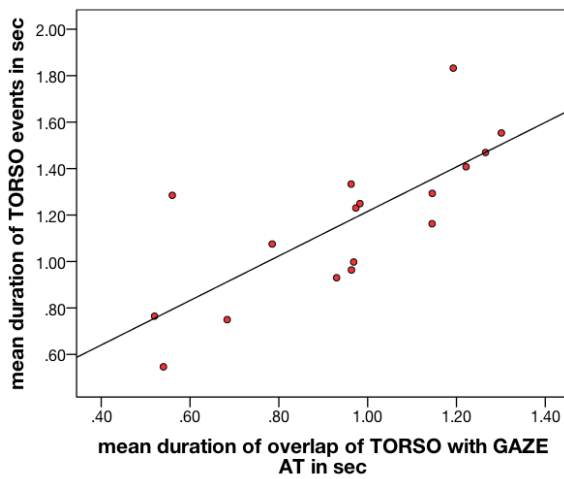
(a) Torso vs. Torso-infant gaze at
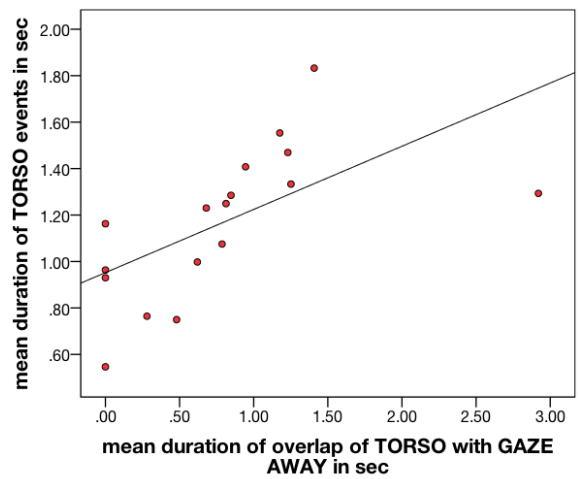
(b) Torso vs. Torso-infant gaze away

Figure 7.7.: Correlation between mean duration of events of infant gaze overlapping with the torso modality and mean duration of events in the torso modality
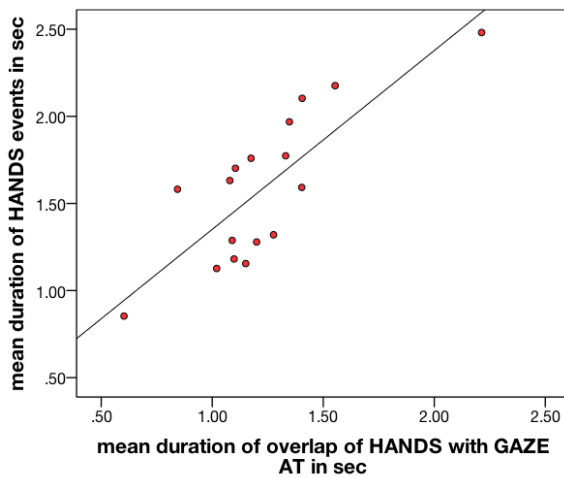
(a) Hands vs. Hands-infant gaze at
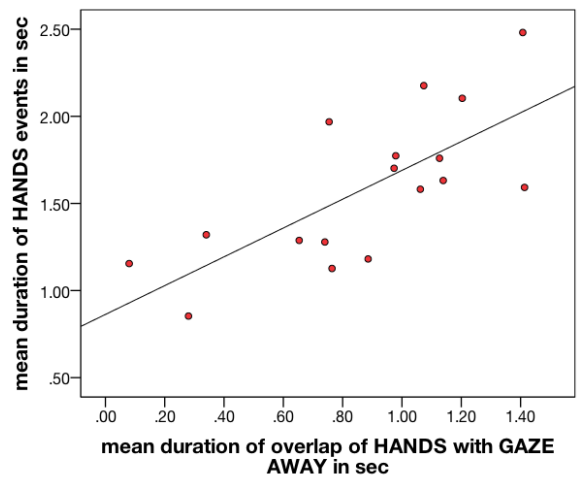
(b) Hands vs. Hands-infant gaze away

Figure 7.8.: Correlation between mean duration of events of infant gaze overlapping with the hands modality and mean duration of events in the hands modality

228

7.4. Summary of findings

To capture the ways in which the multimodal input provided to young infants during natural interactions is coupled to their own behavior, the temporal overlap between classes of maternal vocal-motor synchrony and the infants' gaze behavior was calculated. The data suggest that mothers synchronized their vocal behavior with body movement significantly more when the infants were looking at them than when their infants were looking away. Further analysis of the duration of infant gaze events and the duration of maternal synchrony events showed a strong correlation between them for all modalities. This suggests that the longer the infants spent looking at the synchrony, the longer the synchrony provided by the mother, or to put it the other way round the longer the synchrony the longer the infants attended to it. This suggests that the infants might be attracted to the multimodality. The significant correlation between gaze away and the modalities hand and torso could be explained by the fact that the mothers could be sustaining the synchrony to match their infants' gaze at their hands, or the area of their torso, or event gazes at objects they are holding.

8. Intermodal synchrony and language development

8.1. Objective

The main question guiding this section is whether the temporal synchrony between action and language, as a form of caregivers' responsiveness, in early interaction behavior shapes children's language learning.

In Chapter 4 it was shown that infants' and mothers' gaze patterns change along with the changing capacities of the infants' gaze behavior. Also, in Chapter 5 it was shown that mothers use multimodal practices to recruit and maintain infant attention, thus potentially paving the way to the development of joint attention. In Chapter 6 it was suggested that this tight temporal relationship between mothers' vocalizations and actions, might itself 'package' actions acoustically and make language perceivable and tangible to the infants. The suggestion here is that the congruent overlap of language and action assists infants in discovering that sounds relate, or refer in some way, to parts of events and objects, thus creating a first form of reference. However, in Chapter 7 it was shown that caregivers use significantly more synchrony when acting within the focus of their infant's gaze, and that they also sustain the synchrony for the duration of its overlap with their infant's gaze.

Following up on these analyses the question arose whether these practices–and the variability between subjects–could predict the infants' language development. Following questions were

investigated:

- To what extent is synchrony employed as a form of responsiveness to infants' gaze? This question was partly addressed in the previous chapter. In this analysis the data included the interactions with the six-month-old infants.

- To what extent do infants attend to the synchrony provided in the input?

- Is the variability is this behavior of mother and infant predictive of the infant's later language development?

8.2. Method

8.2.1. Participants and data

Fourteen out of the seventeen mothers agreed to participate in this follow-up study. This sample included nine boys and five girls. Ten of the children were firstborn. The average age of all participating children was 15 weeks (SD = 1) at the first visit and 27 weeks (SD = 1) at the second visit. The average age of the mothers was 32 years (SD = 3.8); they had all completed secondary education and were native speakers of German.

For this analysis the data originated from the same corpus described in the previous chapters. The data from the first and fourth visit were used, i.e., when the infants were three and six months old, respectively. Additionally, data were collected with the ELFRA-2 (Grimm & Doil, 2000), a parental questionnaire (German adaptation of the Infant Form of the MacArthur-Bates Communicative Development Inventories) at the age of 24 months, in order to assess the language development of the infants. For the parental language survey (ELFRA-2), mothers were contacted when their children were 24 months old. They were instructed how to fill out the survey by phone. The survey was then sent to the participants via mail. Their written reports were sent back after several days. For their effort, mothers were reimbursed with a book for their children.

The hypothesis was that the packaging of actions with vocal behavior will have an effect on the infants' later vocabulary. Especially, the prediction was that since acoustic packaging was associated with event processing in younger infants (Brand & Tapscott, 2007), there should be relationship between the maternal behavior and verbs, as words for events.

From the questionnaire following data were used:

a) the overall productive vocabulary

b) the proportion of spoken verbs, and

c) the proportion of spoken nouns

In the analysis of reported verb production, the focus was on two sections of the checklist including 14 modal verbs types and 31 event verbs types. The majority of the event verbs were dynamic.

8.2.2. Coding

For the data on language-action-synchrony, the data reported in Chapter 6 were used. The coded multimodal intervals produced the basic units of analysis, which was named *maternal synchrony*. Furthermore, the gaze data reported in Chapter 4 were used. From the overlap between infants' gaze at the mother and the mothers' synchrony intervals, new measures were calculated. To control for the possibility that all measures depend on the mother's verbal behavior, *maternal talkativeness* was also coded. For this, the Praat data reported in Chapter 6 were used. This resulted in the development of following measures:

- *Maternal synchrony* is the overall duration in seconds of synchrony events (i.e., the maternal behavior in which language was emphasized through bodily movement) divided by the length of the video in seconds.

- *Maternal talkativeness* is the duration of maternal verbal behavior as a proportion over the interaction time.

- *Attended synchrony* represents the proportion of maternal synchrony overlapping with the infant's gaze towards the mother (i.e., the duration in seconds of overlap of maternal synchrony with infant gaze was divided by the total duration in seconds of maternal synchrony).

- *Infant's attention to synchrony*, represents the proportion of the infant's gaze towards the mother overlapping with maternal synchrony (i.e., the duration in seconds of the overlap of infant gaze with maternal synchrony was divided by the total duration in seconds of infant gaze at the mother).
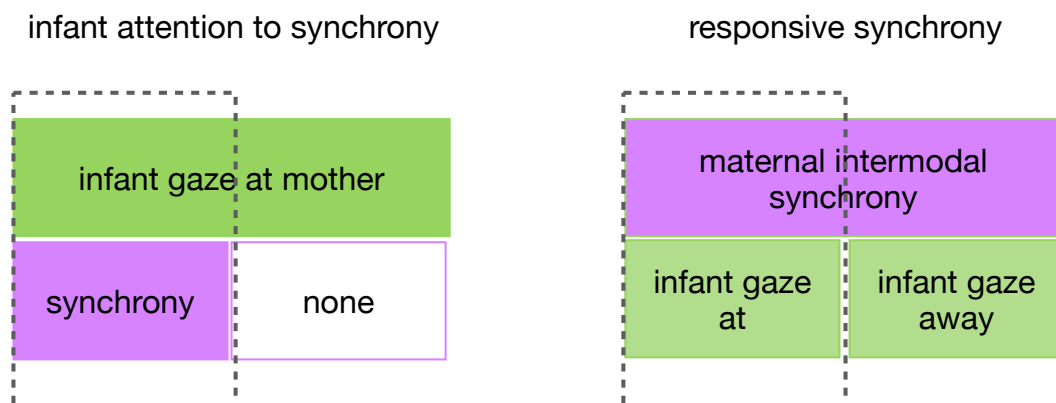
Figure 8.1.: Schematic illustration of measures

8.3. Results

As can be seen in the respective Appendix Tables (D.1 and D.2), there was great variability between subjects in the extent to which maternal synchrony is responsive to the infants' gaze with values ranging from 20% to 98% for the three month olds and from 0% to 76% for the six

month olds. Also, there was variability between subjects in the extent to which infants attend to the synchrony provided by the mother, with values ranging from 13% to 57% for the three month olds and from 0% to 60% for the six month olds. Finally, as presented in Appendix Table D.3 there was great variation in the child's reported vocabulary. There was a child who spoke only 16 of the words in the questionnaire, while at the same time another spoke 215. The same was observed for the proportion of verbs. In this case there was a child who didn't speak any verbs, while at the same time for another child verbs took up one fifth of their overall vocabulary. Therefore, correlational analyses were conducted among the dependent measures to assess the relationships.

8.3.1. Early interaction at three months of age and child's language learning

For the three month-old infants, infants' attention to synchrony was significantly correlated with the overall reported productive vocabulary ($r = .525$, $p < .05$), and the proportion of reported spoken verbs ($r = .530$, $p < .05$) at the age of 24 months. This finding suggests a relationship between the amount of synchrony that the children attend to in early interaction and their later vocabulary skills: the more the infants attended to the synchrony between their mother's action and language when they were three months old, the better the infants' reported overall productive vocabulary and ability to produce verbs at the age of 24 months.

8.3.2. Early interaction at six months of age and child's language learning

At the time when the infants were six months old, providing only synchronized action and language behavior was no longer sufficient, as shown by the lack of a significant correlation between the amount of maternal synchrony and children's reported verb production ($r = .418$, $p = .137$) or overall reported production ($r = .463$, $p = .96$) at 24 months. Instead, there is a significant correlation between maternal responsive synchrony and the children's productive vocabulary ($r = .619$, $p < .05$) as well as the overall proportion of reported spoken verbs ($r = .552$, $p < .05$).

8.3.3. "Switching roles"-model

The two variables, infants' attention to synchrony at three months and mother's responsive synchrony when the infants were six months old, are not independent from each other. The correlation between the two was highly significant ($r = .80$, $p < .001$), suggesting that the maternal behavior when the child was three months old related strongly to her behavior when the child was aged six months. Therefore, the independent contribution of each of the predictor variables on the dependent variable (ELFRA-2) cannot be assessed, due to multicollinearity. For this reason, a linear regression was calculated to get insights into the relation of these two variables with later language acquisition skills. The analysis showed that when considered independently from each other, infant attention to synchrony at three months explained 28% of the ELFRA-2 variance ($F(1,12) = 4.56$, $p < 0.05$), and maternal responsive synchrony at six months explained 38% ($F(1,12) = 7.45$, $p = 0.05$) of the ELFRA-2, and both variables were predictors for the overall vocabulary at 24 months. Also, both infant attention to synchrony at three months (28%, $F(1,12) = 4.69$, $p = 0.05$) and maternal responsive synchrony at six months (30%, $F(1,12) = 5.25$, $p = 0.04$) significantly predicted the variance of the proportion of spoken verbs at 24 months. This analysis confirms the idea suggested by the correlation analyses, that the amount of maternal temporal coordination of action with speech perceived by the infant in early interaction when the child is three months old, and the extent to which mothers provide this temporal synchronization in accordance with the child's attention later, when the child is six months old, are associated with later language development.

8.3.4. Talkativeness

One could argue that it is not the synchrony, as a form of responsiveness, but rather the verbal behavior of the mother that is positively related to the child's vocabulary development. To investigate this possibility, the talkativeness of the mothers was calculated as a proportion of the interaction time that they devoted to verbal activities. At both ages talkativeness of the mother

did not relate to the child's reported vocabulary (at three months: *r* = .209, *p* = .474 ; at six months: *r* = .433, *p* = .122). The same was found for the proportion of spoken verbs (at three months: *r* = .320, *p* = .264 ; at six months: *r* = .376, *p* = .185). This finding is supportive of the model proposing that multimodal behavior—rather than only verbal input—is predictive of language development.

8.3.5. Reported lexicon

To gain a more detailed picture of the child's reported vocabulary skills, different variables from the reported lexicon were analyzed. There was a highly significant correlation between the proportion of spoken verbs with the overall productive vocabulary (*r* = .926 , *p* < .001). This suggests that the more verbs the children were reported to produce, the more extended their overall vocabulary. Interestingly, the proportion of spoken verbs correlated negatively with the proportion of spoken nouns (*r* = -.938 , *p* < .001), and the proportion of nouns also correlated negatively with the overall productive vocabulary (*r* = -.893 , *p* < .001). The finding suggests that in this sample, verbs could be seen as an indicator of overall growing productive vocabulary. This is consistent with the view that verb learning is fundamental to language development (Göksun, Hirsh-Pasek, & Golinkoff, 2010).

8.3.6. Switching roles-model applied to children with low vocabulary

In the analysis of individual differences in language acquisition, particularly two children seem to be at risk for developmental delay, as their reported values for productive vocabulary are the lowest: subject 09 and 13 in Appendix Table D.3. However, since the mother of subject 13 is not speaking during the interaction with her six-month-old child, only subject 09 will be described in more detail. For the fine-grained analysis, a median split of the whole sample was conducted for the critical value of maternal responsive synchrony at thee months of child's age (*Md* = 36,67) and at six months (*Md* = 42,01) of age. For subject 09, both measures were

below the threshold (see Appendix Tables D.1 and D.2, respectively). This finding adds to the results suggesting that maternal behavior in early interaction is related to subsequent productive language development.

8.4. Comparison of language testing tools

To assess language performance, in addition to the ELFRA-2, a standardized test –SETK-2 (Grimm, Aktas, & Frevert, 2000) – was conducted at the families' homes when the children were 24 months old. This test was used to test the language skills of the children with a tool other than parental reports. It consists of two parts: a comprehension and a production part. During the comprehension part the children are shown cards with four pictures on them and are asked to show the correct card for the word or the phrase they hear. In the production part, children are shown one picture and are asked to name the word or phrase depicted in the picture.

For this test, the families were contacted separately. Out of seventeen families thirteen families agreed to participate. Unfortunately, the test could be carried out only with 10 families, as two of the children did not cooperate and in one case the testing could not be completed because the infant was ill.

These two measurements, the reported vocabulary and the experimentally tested language performance, were correlated. Due to missing questionnaire data from one family one further child was excluded from the analysis. Thus, the correlation was carried out on a sample of nine children.

The SETK production scores correlated significantly with the ELFRA-2 production (Kendall's tau τ = .817; p = .002) and the ELFRA-2 verb production (Kendall's tau τ = .761; p = .005). The SETK verb production scores also correlated significantly with the ELFRA-2 production (Kendall's tau τ = .648; p = .016) and the ELFRA-2 verb production (Kendall's tau τ = .592; p = .028). Finally, SETK production correlated significantly with SETK verb production scores (Kendall's tau τ = .857; p = .002).

8.5. Summary of findings

For this analysis the data reported in the previous chapters on maternal synchrony and infant gaze were used and new measures were developed. *Infant's attention to synchrony*, represents the proportion of the infant's gaze towards the mother overlapping with maternal synchrony, *maternal synchrony* is the proportion of the interaction time the mother uses synchrony, *attended synchrony* represents the proportion of maternal synchrony overlapping with the infant's gaze towards the mother. Also, *maternal talkativeness* was calculated as the proportion of the interaction time mothers engaged in verbal practices. Furthermore, when the children were 24 months old, the mothers completed the ELFRA-2 language survey. Out of this questionnaire the overall productive vocabulary was used. Also, the proportion of spoken verbs, and the proportion of spoken nouns were calculated. Finally, a subgroup of the original sample agreed on participating in a standardized speech evaluation test which provided real comprehension and production data apart from the parental report.

The synchrony provided by mothers as an individual measurement was not found to be associated with language development. The same is valid for maternal talkativeness. It did not correlate with the language development measures. The correlation analyses and subsequent linear regression revealed that infants who at three months of age looked at more synchrony provided by their mothers could speak more verbs and had overall larger vocabularies at 24 months. Also, mothers who were more responsive to the gaze of their six month-old infant when providing their synchrony had toddlers who could speak more verbs and had overall larger vocabularies at 24 months. Finally, infants who spoke more verbs - and not more nouns - showed an overall more advanced language production.

Also, in the comparison of the language assessment tools the scores correlated significantly with each other, suggesting that the maternal reports did indeed reflect the language skills of the children. One drawback was that in some cases the test proved impossible to be carried out, as the children did not cooperate, thus reducing the sample size on which the comparison could be made.

9. Discussion

The core assumption of the present study was that in early interactions infants' perception is guided, and thus educated, towards certain aspects of the world which caregivers consider relevant for language learning. This education of attention was assumed to be both towards picking up perceptual cues, such as sights and sounds belonging together or perceptually salient events, as towards perceiving social cues, such as eye gaze and gestures. Most importantly, the present study proposed that development is guided by processes not residing *within the infant*, but by processes constructed *within interaction*.

For the above-defined education of attention, the argument raised was that the development of perceptual and social attentional skills is supported through reinforcement from the environment. If so, then one would expect to find some evidence of the influence of mothers on infants. This was formulated in the first hypothesis:

H1: Infants' attention will be educated by the interaction. More specifically, there is a relationship between infant gaze patterns and maternal gaze patterns. Also, there should be evidence of the development of this relationship as a result of the interaction history of the dyad.

Here, infant attention was operationalized as the infant's gaze at the mother's face. The first prediction was that there should be a relationship between infants' gaze at the mother's face and mothers' gaze at the infant's face. Furthermore, the interactional history of the dyad was operationalized as the comparison of mother-infant gaze patterns at two time points. The second

prediction was that if mutual attention is affected by the interaction then there should be evidence of a convergence of mother-infant gaze behavior over time. As a result of this convergence, mother and infant should develop mutual gaze patterns. Mutual attention was operationalized as the overlap between mothers' gaze at the infants' face and infants' gaze at the mother face. This hypothesis was addressed with the analysis presented in Chapter 4 (Mother-infant gaze patterns: a longitudinal comparison). More specifically, in this section the gaze patterns of mother and infant were investigated and the question whether these patterns change over time.

Qualitative analyses revealed that when the infants were three months old they showed a main interest in the mother's face. They gazed at the mother's face for prolonged periods, following her face as she moved during the diapering activity. While the mothers actively switched their gaze to the infant's face and away from it, the infants did not switch their gaze often.

Analysis of the interactions when the infants were six months old showed that the infants gazed at the mother's face only for short intervals and some of the mother's gazes at the infant were completely unnoticed by the infant. The analysis made the changing interest of the infant apparent. Although infants still showed interest in the mothers' face, they did not engage in prolonged episodes of eye contact. There was an increased interest in exploring the environment and exploring the objects involved in the diaper changing activity. The mother adapted to this change by actively seeking to engage in eye contact with the infant, by following the infant's interest in objects, and making the objects available to the infant for visual and tactile exploration. The examples presented in the analysis spoke for a change in the gaze patterns of both mother and infant over time.

Quantitative analyses of the interactions with the three-month-old infants showed that mother and infant spent much time in mutual attention. During one third of the interaction time infants and mothers gazed at each other. The mother gazed at the infant's face the most. This was the same for the infants. This preference for looking at the other's face was also reflected in the duration of gaze events, which were the longest for gazing at the other's face. This behavior was followed by gazing at objects or the infant's body. The three-month-olds did not gaze at

objects that often. Finally, mothers spent very little time gazing away. They focused on the task at hand and the interaction with the infant. Infants, on the other hand did spend time looking away, exploring the room and the surroundings, although this did not occur very often. For the six-month-old infants mutual attention occurred during one fifth of the interaction time. Again, gazing at each other's face was the most frequent gaze behavior, but the infants also showed an increase in gazing at objects and gazing at the surroundings. Also, at six months infants spent longer intervals when gazing at objects, and the duration of gazes at the mother's face were reduced.

A statistical comparison of maternal and infant gaze at the two time points revealed a decrease in looking at the partner's face. However, this was statistically significant only for infants. This suggests that the mothers maintain their gazing behavior. Yet, at the age of six months the duration of gaze intervals becomes significantly shorter for both mother and infant. For the infants, Messer and Vietze (1984) suggest that this could be due to the fact that, with increasing age, infants become more efficient in using shorter glances to obtain information. Holmlund (1995) also presents similar findings for the development of turn-taking in the three first months. This change of the infants' gaze pattern alters the social interaction, which would provide an explanation for the development of the mothers' behavior. Infants become faster and mothers adapt by becoming faster in gaze switches to and from the infant's face. This is consistent with Kaye and Fogel (1980) longitudinal observations of 6-26-week-old infants, according to which mothers' adjusted their behavior to the attentional preferences of the infant. They found that mothers modified the timing of their expressive behaviors. This could represent the adaptive maternal behavior found in the microdynamics of gaze organization: As the infant becomes more skilled and faster in controlling his or her gaze, the mothers adapt their gaze behavior to that of the infant. This is consistent with Fogel and Garvey (2007) "alive communication", i.e., the mutual adjustment to create a coordinated joint activity.

Moreover, eye contact was reduced significantly both in terms of the proportion of the interaction time as well as duration of eye contact intervals. This shift of interest is consistent with an

increasing interest in objects, which was the result of the statistical comparison of the two points of observation. At six months infants–but not mothers–gazed significantly more at objects and away, but the duration of gaze intervals did not show any significant change. This is consistent with studies (Bakeman & Adamson, 1984) suggesting that at about the age of six months, infants are intensely interested in the exploration of objects, yet attend to single aspects, such as either the face or the objects. The significant decrease in the duration of gazes at the mother's face in combination with the lack of such a significant change in the duration of gaze at objects further support the view that, at six months infants are disengaging from the face-to-face orientation and learning to engage in object focused attention. This finding is certainly in line with the first hypothesis formulated above. It provides evidence for a development of gaze patterns as a result of the history of the interaction.

Further support of the influence of interaction history is provided by the correlation of mothers and infants' gaze at three and six months. This analysis revealed that mothers retain their gaze profile in time: Those mothers who looked a lot at their infants' face when the infants were three months old, also looked a lot at them at six months. This was not the case for the infants, though. It seems that infants undergo some development between the third and sixth month and they modified their gazing behavior. This, again, is consistent with Messer and Vietze (1984) who suggest that the development of gaze may be due to the acquisition of new methods of communication, such as gestures and intentional vocalizations (in their research at 26 weeks). Yet most importantly, when looking at the cross correlation of the gaze data at the two time points an interesting pattern arises. Although at three months the proportion of the interaction time which infants spend looking at their mother does not correlate with the mothers' gaze at the infants, three months later–at six months–the gaze behavior of mother and infant does correlate significantly. This indicates that mutual attention undergoes a development in that it is established as a method of communication. Mother and infant by six months coordinate their gaze behavior to each other such, that those mothers who looked at their infants' face more had infants who also looked at their face more. This clearly is evidence of a relationship between

infant gaze and maternal gaze, as predicted by the first hypothesis. The fact that at three months, this pattern is not yet observable, but it appears at six months provides also support in favor of the second prediction made by the first hypothesis, namely that there should be evidence of some kind of convergence of mother-infant gaze behavior over time.

Finally, in Chapter 4 a further analysis focused on the role of mother and infant in structuring eye contact. The observation was that between the two time points there was a role reversal in gaze behavior. For the three-month-olds, it was the infant who was looking at the mother's face for long periods, whereas it was the mother who set the pace for the organization of eye contact; for the six-month-olds, it was the mother who framed the gaze episodes, as the infants switched their gaze to and away from the mother, more actively. Evidence confirming this observation would also support the hypothesis that mother and infant influence each other's behavior, as it would suggest that they develop their own gaze organization, as a result of adapting to each other's behavior over time. In line with the hypothesis, the statistical analysis revealed that mothers ended the eye contact significantly more frequently when the infants were three months old than when they were six months old. In contrast, infants terminated eye contact significantly more frequently when they were six months old than when they were three months old. Also, at three months infants framed mother's gaze more than expected by chance, and respectively, at six months mothers framed infant's gaze more than expected by chance.

This finding not only supports the hypothesis outlined above, but it suggests that mother and infant develop their own patterns of focusing attention on each other, by adapting to, or co-regulating with (Fogel, 1993a), each other's behavior. As the infant matures, his or her behavior changes as a result of the interaction bringing about also a development in the behavior of the mother. Thus, when considering the development of attention, it seems reasonable to focus on both mother and infant and their contributions to the interaction. To this end, the analysis discussed so far revealed that mothers actively structure their behavior, in this case their gaze, and adapt to the behavior of the infant. This observation is in line with a further argument put forward in the introduction, according to which caregivers use infants' predispositions upon

which they 'hook' their behavior. The suggestion was that caregivers intuitively act in a way that fits the infants' own mechanisms of development. This is consistent with Zukow-Goldring (1997) work, who proposes that caregivers make affordances and effectivities perceivable to the infants. Accordingly, the assumption then was that the education of attention is the result of the active structuring of the input in systematic ways. What is meant by active structuring is that the information provided to infants is selected to be in accordance with their mechanisms of perceptual development. This is consistent with Gogate et al. (2001) who showed a reciprocity between infants' perception and mothers' provision of various types of –in their case audiovisual–communication.

One mechanism suggested as driving infants' perceptual development was the symbiosis of language and action (Zukow-Goldring, 1997; Hirsh-Pasek & Golinkoff, 1996). Accordingly, early in development infants selectively attend to intermodally specified events presented redundantly. This redundancy is deemed necessary early in development as it enhances perceptual processing, learning, and eventually memory for multimodally specified properties (Bahrick & Lickliter, 2000; Bahrick et al., 2004). The most important perceptual attribute was discussed as being the temporal synchrony with which the temporally overlapping intersensory information is presented (Lewkowicz & Kraebel, 2004; Lewkowicz, 2000). If, according to interactionist approaches, mothers actively structure their behavior so as to match the developmental stage of the infant, then one would expect to see multimodal structuring of interactions by the mothers, as this would assist young infants in perceiving the maternal behavior as salient and focusing their attention to it. The second hypothesis put forward was, thus:

> **H2: Even in early interactions, mothers will use temporal synchrony to educate**
> **infant attention. There will be a systematic structuring of language and action**
> **in multiple modalities.**

Firstly, synchrony was operationalized as temporally coordinated maternal vocal behavior overlapping with body movements in various parts of the mother's body. According to the

assumptions presented on the role of intersensory redundancy for perceptual development, the prediction here was that, for the younger infants, a significant part of the interaction with their mother should be structured in a multimodal way. This is because intersensory redundancy is considered most important early in infancy. The second prediction was that at the second point of observation the vocal-motor synchrony in maternal stimulation will increase. This was expected as at six months of age, infants will have become more responsive to social signals and, thus, to maternal input (Trevarthen, 1979; Lavelli & Fogel, 2005). This constantly increasing, interactive mutuality, would provide positive feedback for the mother. The mother would then, in turn, increase stimulation in a attempt to prolong the duration of the sequences of positive affect and adapt to the infant's developing interaction cycles.

Secondly, systematic structuring referred to the way in which verbal behavior reinforces action and the way in which bodily action reinforces the verbal signal. This is based on the expectation that the interplay of language and action, on the one hand helps infants create a link between what words mean and what they refer to in the world, thus educating their attention for relevant aspects of speech. On the other hand, the symbiosis of language and action helps infants parse the world into meaningful events, thus educating their attention towards the understanding of actions. This view is reflected in the work of Zukow-Goldring (1997) and her socio-ecological approach to language learning and the work of Hirsh-Pasek and Golinkoff (1996) and their notion of acoustic packaging. Here, the prediction was that even in the first half of the infants' first year the input provided by the mothers would be rich in such structure. Furthermore, the prediction was that between the first and second point of observation there should be a difference in the structuring provided by the mother. As the infant moves from the interpersonal to the object-focused stage (Bakeman & Adamson, 1984; Stern, 1974) one would expect this structuring to change so as to correspond to the infants' perceptual development. This change would then be apparent in the modalities used to structure the input.

This hypothesis was addressed in the analysis presented in Chapter 6 (Synchrony between action and language). Accordingly, it was investigated whether the spontaneous vocal input to

three and six-month-olds contains action information provided at the same time. Qualitative analysis sought to reveal, during which practices and in what forms, the co-occurrence between language and action is achieved. Five classes of bodily movements that were synchronized with vocal activities were identified. Also five vocal types were identified that were used during these multimodal practices. Using these coding categories, various multimodal practices were identified in which action (in the form of bodily movements) and language (in the form of different vocal types) temporally overlapped each other. Looking at the data qualitatively, there was variability in the way language and bodily movement can temporally overlap. The observations suggest that the temporal relationship between two modalities is a complex phenomenon — as Zukow-Goldring and Rader (2005) have already pointed out in this context. Specifically, synchrony does not always refer to simultaneous occurrence in time. Inspired by the basic interval relations (Nebel & Bürckert, 1995; Kopp & Bergmann, 2010), the observation was that two intervals – in this case the oral and motor modalities – overlap in various ways in the data. For example, the vocal signal can introduce or announce the action, that is, occur earlier than the actual action. Also, language can finish or comment on the action performed, that is, occur after the referred action. Furthermore, an overlap could occur at the beginning of an action or vocal behavior, communicating to the infant that something new is being initiated. Overlap could also occur at the end of an action marking the goal. Complementarily, an action might be used to highlight the end of an utterance, signifying to the infant that it is his or her turn to speak. Finally, one element during an action or verbal phrase may be especially highlighted. The suggestion here is that the notion of synchrony should be broadened to cover such partial overlaps, because the various forms of co-occurrence might actually provide a differential structuring, or packaging, of actions or language.

These findings support the hypothesis formulated above about the systematic structuring of the input. They show that, even in very early interactions with three- and six-month-old infants, mothers provide input that is very rich in intersensory structure. Analysis revealed that vocal and motor modalities can reinforce each other but also structure each other: Regarding

the reinforcing function, bodily movement can emphasize specific aspects of speech, making the reference salient and providing a non-arbitrary link between what words mean and what they refer to in the world (Zukow-Goldring, 1996; Stern et al., 1977; Gogate et al., 2000; Gogate & Bahrick, 2001; Rolf et al., 2009). The structuring function could be demonstrated in body movements 'packaging' phrases; thus, the structuring function provides packages of actions, because language can link or 'chunk' bodily movements by describing them, introducing them, highlighting their goals, or segmenting them into smaller action steps that eventually facilitate learning (Hirsh-Pasek & Golinkoff, 1996). Brand and Tapscott (2007) did, indeed, confirm that infants considered sequences of actions as belonging together, when they were "packaged" by narration. Consistent with this work, the present findings are interesting because they define a path of meaning development in children that narrows the possibility of a prelinguistic period, as language functions rather as a social signal that packages motion into meaningful events and can help shape perception early in development (Majid, Bowerman, Staden, & Boster, 2007; Majid, Boster, & Bowerman, 2008).

In addition to the qualitative results, data were analyzed quantitatively, because of the interest in the temporal extent to which mothers offer a multimodal stimulation during the everyday interaction, and which different modalities overlap with language. Analysis showed that vocal signals were used in 39% of the interaction time when the infants were three months old and 36% of the time when the infants were six months old, suggesting that they play an important role in early interaction with infants.

Furthermore, when the infants were three months old 73%, and at six months 70%, of these vocal signals occurred in synchrony with movement of various parts of the body, suggesting that infants experience language multimodally. Thus, the data reveal that a considerable part of maternal vocal stimulation is language accompanied by synchronous bodily movements. This is in line with the hypothesis formulated above. It suggests that in these early interactions with their infants mothers do make extensive use of intersensory redundancy. While previous research has shown experimentally that infants are sensitive towards intersensory redundancy, the research

presented herein shows for the first time that during everyday activities, mothers actually do provide language in this redundant form.

When comparing the two time points, it could be observed that the overall use of maternal coordinated speech and body movement did not change significantly between the two observation times. However, the modalities that convey the synchrony seem to be age-dependent. More specifically, in comparison to interacting with three-month-olds, when interacting with six-month-olds, mothers (a) significantly increased the use of the facial modality, while significantly reducing the use of the head modality, and (b) significantly increased movements of their upper body (looming in and out). Furthermore, a comparison between the duration of synchrony events at the two observation times revealed a main effect for age, suggesting that as the infant becomes older the synchrony events used by the mother became shorter.

The next analysis step focused on the extent to which infant stimulation involves not only bimodal vocal-motor synchrony, but a redundant stimulation through synchrony across more than two modalities. When testing the influence of age on the occurrence of multiple overlapping modalities, a main effect for age was found for the combination of two and three modalities. As the infants became older the mothers reinforced their verbal behavior in a more overlapping manner.

These findings are interesting with respect to the prediction that there should be an increase of the use of synchrony over time. Even though the overall synchrony does not increase from the first to the second observation point, there is an increase in specific modalities. For example, there is an increase of the facial modality and the torso modality. The facial modality is a modality which conveys social feedback, i.e., a smile, a frown, or a surprised face. The torso modality allows the mother to reduce and increase the distance of her face from the infant's face. This is in line with the prediction outlined above, which stated that as the infants become more responsive to the input, the mothers will provide more multimodal input. The results suggest that mothers do provide more multimodal input, but only at those modalities which convey conventionalized social feedback. Also, the fact the the synchrony events become shorter speaks

in favor of an intensification of the multimodal input, suggesting that it becomes faster and increases in frequency. Finally, the finding that at six months of infant age the mothers provide a multimodal input with increased overlap of multiple modalities of vocal-motor synchrony is also in line with the hypothesis that with increasing age the multimodality increases.

The analysis further revealed a relationship between certain types of vocal behavior and certain body types. Mothers were more likely to use emotionalization signals such as sighing, exaggerated breathing, or panting when engaging in facial movements. By using these ostensive cues, mothers may be attempting to attract infant attention; Senju and Csibra (2008) demonstrated convincingly that infants are sensitive towards ostensive cues. In the data presented above, mothers used this combination at the beginning of new actions or utterances recruiting the infant's gaze and communicating to him or her that something interesting is about to happen. A further relationship was found between the imitation of infant vocalization and both facial expressions and head movements. When mirroring the behavior of their infant, mothers evidently exaggerated their behavior by using movement of the head and the face. M. Papoušek and Papoušek (1989, p. 159) explain the use of imitation of infant vocalization in terms of mothers "provid[ing] models which may most easily elicit matching responses". In the data, when mothers imitated the infant, they were most likely to use movement of their head and their face, maybe to further motivate the infant to participate in the interaction and repeat her vocalization. Finally, a relationship was found between speech and hand movement. This is interesting considering that there are almost no conventionalized gestures in the data. Maybe because of the nature of the setting, the mothers may have been using manipulative actions instead of co-expressive gestures, projecting their co-expressive behavior on to the body of the infant and the objects available in the immediate environment giving, thus, a tactile sensation to the vocal signal. These relationships were found for both observation points. These final findings provide further evidence of the hypothesized rich structuring of language and action in these early interactions.

Looking at the results discussed above concerning the use of synchrony in combination with the results discussed for the development of gaze patterns suggests that mothers might modify

their multimodal behavior to match the developing gaze patterns of their infant. The fact that gazes at each other become significantly shorter is in concert with the present finding that synchrony events also become shorter. Furthermore, the idea presented above that infants gaze less at the mother because they become more efficient is also in concert with the finding that the mothers reduce head movements and increase face movements. Mothers may be keeping track of their infants' increasing perceptual abilities and do not need to use big movements such as head nods; instead, they provide more subtle conventionalized social signals such as smiles, frowns, and eyebrow movements. This finding echoes the work of Bruner (1983) who sees the development of the infant as a system which moves from being responsive to biological signals to becoming sensitive to individual differences and cultural practices. In this case, the behavior of the mother could be interpreted as scaffolding the way to conventionalized forms of communication. Nevertheless, the important aspect here is that this multimodal input is not just a behavioral modification activated when interacting with the infant, but that it is selectively used in a collaborative co-construction of the input to educate infant attention.

Regarding the mechanism enabling the education of attention through synchrony, the argument raised in the introduction was that, in early mother-infant interaction, mothers will selectively reinforce mutual attention. The reinforcement was suggested as being a co-occurring behavioral modification. This would educate infants to engage in mutual attention as it creates the expectation that when they do something different happens than when they do not. The behavioral modification proposed was intermodal synchrony provided by the mother. This multimodal input could frame episodes of attention, making them salient for the infant and making them rewarding for attention and memory. The framing in this case would be the provision of specific conditions, which prevail at the moment episodes of eye contact occur. According to Bruner (1983), this would create a context for this behavior making it recognizable and anticipated. At a perceptual level, language and action provided in synchrony could reinforce each other, attracting and educating the infants' attention. At a social attentional level, through repeated interactions within everyday routines, certain signals, in this case mother-infant mutual

attention, or infant gaze at the mother, could become established as relevant and anticipated by the infants, who would be thus socialized into perceiving them and using them. The third hypothesis formulated in the introduction was:

H3: Mothers will use synchrony to reinforce infant gaze. The employment of synchrony will, thus, be coupled to the gaze behavior of the infant.

This hypothesis was addressed in Chapter 5 (Educating Attention within multimodal interactions) and Chapter 7 (Responsive Synchrony). The data analysis in Chapter 5 addressed the question whether, and how, the development of mutual attention is supported and, thus, educated in early interactions through social reinforcement. For this purpose, the analysis aimed to explore whether eye contact, as one early instantiation of mutual attention, is reinforced behaviorally in natural mother–infant interactions, and how this goes hand in hand with the development of the infant and changes as a function of the history of the joint interaction (Flom et al., 2004). The focus was on the sequential organization of gaze behavior in search for maternal strategies in recruiting, maintaining, and pursuing eye contact with her infant.

The analysis showed that episodes of eye contact were systematically reinforced by accompanying verbal and nonverbal behaviors that—according to the intersensory redundancy hypothesis (Bahrick et al., 2004)—can be viewed as a form of attentional frame that makes episodes of eye contact an exciting and rewarding experience. More specifically, the analysis revealed that eye contact episodes overlapped with verbal behaviors that included opening sequences such as greetings, questions, or calling the infant's name—behaviors that have been suggested to have ostensive power, because they signify to the infants that something is explicitly designed *for* them and addressed *to* them. This observation is in line with recent studies testing the effects of such social signals: Senju and Csibra (2008) found that 6-month-old infants followed the adult's gaze toward an object predominantly when it was preceded by ostensive cues, such as direct gaze and greeting in form of infant-directed speech, and not when it was presented with a moving cartoon image or a greeting with intonation of adult-directed speech. The authors argued that

infants do not just react to "nonsocial attention getters" (Senju & Csibra, 2008, p. 670). Consistent with these findings, it was shown that eye contact systematically occurs simultaneously with other social signals. In fact, one can further suggest that infants' sensitivity to ostensive signals may result from a recurring reinforcement of eye contact, because these experiences of multimodal redundant behavior may be rewarding for them. To put it in other words, ostensive signals might be those that are established and learned, because they may be the most reliable in recruiting infants' attention.

Furthermore, the analysis showed that the verbal behavior accompanying eye contact overlapped with not only movements of the mother's body such as facial expressions, head movements, torso and hand movements, but also tactile stimulation of the infant. These multimodal behaviors were provided in a very tight temporal relationship to one another, thus forming multimodal chunks of behavior that frame episodes of eye contact. In these chunks, it was shown that behavior occurred systematically when the eye contact was established and less frequently when eye contact was lost. Therefore, the suggestion made is that through this behavioral modification, which selectively occurs during eye contact and infant-directed speech, interactional formats of mutual engagement (Fogel et al., 2006) in eye contact may emerge, evoking in infants an expectation of rich sensory, rewarding experiences (Bahrick et al., 2004). This is based on the finding discussed above that multimodal information is particularly powerful in capturing the attention of the infant, because it is matched to the infant's mechanisms of perceptual learning (Bahrick et al., 2004). To maintain attention, mothers used repetition of the multimodal packages of stimulation, which they clearly separate from each other with pauses. By repeating their behavior, mothers are producing a predictable behavior. This may help to stabilize the format of participation in eye contact, by entraining infants in prolonged engagement as they become capable of forming and confirming their predictions about the interaction. Finally, to pursue eye contact it was shown how multiple modalities are mobilized so as to elicit a specific behavior from the infant. The pursuit of a particular response has been suggested to be an educational tool used by parents in conversations with children learning how to speak. It has been studied

as a strategy to attention to an absent or inadequate response of the child (Filipi, 2013). In the analysis presented above, it was shown how it is used in the process of teaching an infant to participate in episodes of mutual attention.

In addition, the analysis of interaction sequences revealed that to recruit attention, the multi-modal input provided to the infants during eye contact is triggered interactively. Thus, the infant takes an active role in co-constructing the ostensive input that she or he receives, because the mother uses the focus of the infant's attention as a *'hook'* for her contingent behavior. The effect of contingent behavior has been demonstrated in many studies (Kärtner et al., 2010; Striano & Reid, 2006). The proposal made herein is that it may further reinforce the education of attention. Also, the pursuit of a particular behavior illustrates the way in which the education of attention is dynamically negotiated. This exemplifies the process of co-shaping mutual events as emerging both during the course of a particular interaction sequence and over the interactional history of mother and infant (Fogel et al., 2006). In the analysis it was shown, that a repertoire of interactional formats is established such that not any response of the infant is treated as relevant, but a specific action sequence is expected for the interactional episode to be successful.

The above formulated hypothesis was also addressed in the analysis presented in Chapter 7 (Responsive Synchrony). The question asked in that section was how the infant–as an active interaction partner–regulates the behavior of the mother, and how the multimodal input is collaboratively constructed within everyday activities. Methodologically, this analysis took the qualitative analysis presented in Chapter 5 further, as it included microanalytical coding and statistical analyses, which generalized the observations to the entire data corpus.

More specifically, the temporal overlap between classes of maternal vocal-motor synchrony and the infants' gaze behavior was calculated. The analysis revealed that mothers provided their multimodal behavior significantly more when the infants were looking at them than when their infants were looking away. Further analysis of the duration of infant gaze events at the mother and the duration of maternal synchrony events, showed a strong correlation between them for all modalities. This suggests that the longer the infants spent looking at the synchrony, the longer the

synchrony provided by the mother was, or to put it the other way round the longer the synchrony the longer the infants attended to it. This finding is in line with the observation made in the qualitative analysis, which showed that mothers modify the modalities they use and position themselves and the objects involved in the interaction according to the infants' direction of gaze. The association found between the duration of infant gaze events and mothers' synchrony events also suggests that the infants might be attracted to the multimodality.

The proposal arising from the current analysis is that, by actively adapting their multimodal practices to the infant's attention, mothers are reinforcing the input, making it more ostensive, more unambiguously addressed to the infant. This may be an early mechanism for the education of infants' engagement and attention, leading to sensitivity for ostension and to joint attention in the months to follow. This possibility stands in contrast to the approach formulated within Natural Pedagogy (Csibra & Gergely, 2009), claiming that infants are evolutionary predisposed to attend to ostensive cues. Instead, the results above suggest that infants become educated towards interactional frames.

The findings presented in support of the above-formulated hypotheses suggest that the gaze behavior and the multimodal input are shaped by the interaction and by the development of the infant. It is, thus, reasonable to suggest that variability in this shaping process would be reflected in the later behavior of the infant. This assumption echoes socio-cultural theory and especially the work of Vygotsky (1978) who suggests that interaction characterizes development prospectively. According to this account, different qualities of mother-infant interactions will lead to different developmental outcomes (Keller & Gauda, 1987; Bornstein & Tamis-Lemonda, 1997). Furthermore, if language acquisition is the emergent result of a coalition of cognitive, attentional, and social factors (Hollich et al., 2000) then it is reasonable to suggest that differences in the interplay of these factors would bring about differences in the way language develops.

Departing from the findings in literature reporting that synchrony between two modal events appears to be a key concept in early social learning (Bahrick et al., 2004; Cangelosi et al., 2010), the aim was to shed more light onto responsiveness as a multimodal phenomenon. Up to date,

little is known about how synchrony between action and language, as a form of responsive behavior (Zukow-Goldring, 1996; Gogate et al., 2000; Nomikou & Rohlfing, 2011; Nomikou, Rohlfing, & Szufnarowska, 2013), is related to later language acquisition. It was argued that providing language as a social signal accompanying actions and events early in infancy might facilitate the infants' action processing; a more elaborated understanding of actions and events involving agents, in turn, might provide a robust conceptual basis for the meaning of action words. Following this logic, following hypothesis was put forward:

H4: The variability in the infant's attention to the mother and in the mother's responsiveness in early interactions will be associated with the variability in the child's language acquisition and, especially, with variability in the acquisition of verbs.

This hypothesis was addressed in Chapter 8 (Intermodal synchrony and language development). Infants' attention to the mother was operationalized as the proportion of infant gaze towards the mother overlapping with maternal synchrony. Mothers' responsiveness was operationalized as the proportion of maternal synchrony overlapping with the infant's gaze towards the mother. Furthermore, as synchrony was hypothesized as involving the structuring of actions, it seems reasonable to assume that the effect of responsive synchrony should especially support the acquisition of words for actions, namely verbs. For this reason, the ELFRA-2 parental report questionnaire was used as a measure of the children's language and verb skills.

For the three-month-olds, there was a positive and significant correlation between the amount of synchrony the children attended to in early interaction and their later vocabulary skills. It should be noted here that, as discussed in the sections on gaze patterns of mothers and infants, at this young age the child is looking at the mother to a greater extent. Hence, interaction has been suggested as being in the "affective reciprocal phase" (Colas, 1999, p. 114) at the age of three months. At this stage, interactions are dyadic and interpersonal, meaning that the main interest lies in coordinating attention within the dyad (Schaffer, 1984). Thus, because the child

focuses on maternal behavior anyway, the positive relation to later language acquisition seems to be fostered by the mother providing synchronized action and language behavior at this age. In other words, at this young age of three months, mothers' behavior in early interaction seems to already foster the child's later language development by providing synchrony between actions and speech.

When the infants were six months old, just providing synchronized action and language behavior no longer seemed to suffice. Instead, a significant correlation was found between maternal responsive synchrony and children's later reported vocabulary. This finding strongly suggests that at an older age it becomes important for the mother to monitor the focus of attention of her child, who is now no longer just looking at her but is also attracted and distracted by events and objects in the surroundings (Bakeman & Adamson, 1984). This is also in line with the 'switching' of their gaze roles between the ages of three and six months discussed above. Thus, while at the age of three months, the child focuses on the mother and attends to her responsive behavior anyway, at the age of six months, it is the mother who needs to actively search for slots of infant attention to her and to make her responsiveness perceivable. In other words, the suggestion here is that to foster the child's language learning when the child is six months old, the mother needs to provide synchrony between her speech and action precisely when the child is attending to her.

Incorporating both the behavior of the mother and the infant during early interaction into two dyadic measures, the predictive role of the temporal coordination between language and action as a form of maternal responsiveness to infant attention was investigated. While in previous research, it has been shown that events packaged in narrations were perceived as one chunk (Brand & Tapscott, 2007), it has not been investigated to what extent early packaging in maternal input could bear a relation to the development of meaning for verbs. On the basis of the results from the whole sample as well as individual children potentially at risk for language delay, a 'switching roles'-model was developed, according to which at three months of age, during interaction, children spent significant time looking at the mother. At six months of age, however, it is the mother who is responsive to the infant attention. The results reveal that three-month-old

infants who perceived more synchrony provided by their mothers could produce more verbs and had overall larger vocabularies at 24 months. Mothers who were more responsive to the gaze of their six-month-old infants when providing their synchrony had toddlers who produce more verbs and had overall larger vocabularies at 24 months than their peers. Finally, infants who produced more verbs – and not more nouns – had overall more advanced language production than their peers.

The suggestion here is that already in early interaction, language functions as a social signal packaging and structuring events for infants' perception (Brand & Tapscott, 2007). Mothers who are responsive to the infants' engagement and provide multimodal language involving sensory rich dynamic events may make meaningful action units salient. This supports the acquisition of verbs. Furthermore, the acquisition of verbs may support infants' capacity of perceiving semantic relationships between objects and participants in events, thus leading to a broader vocabulary altogether.

Furthermore, the analyses did not reveal a relationship between maternal talkativeness during early interaction and the children's later vocabulary development. This further strengthens the argument that it is multimodality that makes the input more rich in these early interactions. It is important to note, however, that the measure for talkativeness encompasses only the quantity of speech. In the literature on responsiveness, in contrast, specific forms of maternal verbal behavior (e.g. semantically contingent talk) have been extensively discussed as facilitating vocabulary growth rather than the amount of speech overall (McGillion et al., 2013). These results thus add to this line of research suggesting that instead of considering only the quantity of provided speech, it is crucial to regard verbal behavior in the context of interaction and (the age of) interlocutors.

Also, the analysis did not reveal a relationship between maternal synchrony alone, i.e., not as a form of responsive behavior, and children's later vocabulary development. This further supports the idea that it is not synchrony as such, but synchrony that is responsive to infant behavior that fosters language acquisition.

Finally, the analysis presented above indicates that maternal responsiveness comprises different aspects of the multimodal input at different ages: At three months of age, it is important for the infant to be exposed to multimodal input for longer, whereas at six months, it seems important that the input is responsive to the infants' attention. In this sense, being responsive and communicating with bodily movement might tune infants to interaction and promote later language skills: "moving with others" can become "speaking with others" (Rączaszek-Leonardi et al., 2013). This is in line with an embodied and distributed view on language (Rączaszek-Leonardi, 2009; Cowley, 2011), according to which early language development progresses from coordination of multimodal behavior on a temporal level to joint actions in form of communicative and meaningful acts (Rączaszek-Leonardi et al., 2013). When learning, children take advantage of interaction that spreads information across learning domains. Furthermore, this analysis contributes to the field of language acquisition research by suggesting that when working with longitudinal data, it is necessary to consider that the measures used for comparing different ages are age-dependent, because different factors or cues might be weighted differently in various stages of development (see also Hollich et al., 2000).

This main assumption presented in the present work follows an interactionist embodied cognition view on language acquisition, suggesting that language emerges out of a synergy of perceptual mechanisms and social interaction (Hollich et al., 2000). Accordingly, infants use various sources of information or cues which guide their learning: from salient perceptual information, to the movement of their body and their interaction with the environment, to their engaging in social interactions with other humans. This grounds language in perceiving the world and (inter)acting with it.

This thesis constitutes a contribution to the investigation of the interaction of the infant with the environment, and more specifically to the question how this interaction can prove helpful for the acquisition of language. The suggestion made is that social interaction is a tuning device for the infants' senses. By participating in interactions the infants' perception is calibrated so as to select, to perceive, and to act on those aspects of the world that are relevant for his or

her language learning. According to this view, theorizing on language development should depart from considering very young infants, long before they utter their first words, focusing on the continuity of development. For this reason, the present study captured infants' interactions starting at three months of age and accompanied their development until they became eight months old. The main interest was in the detailed observation of mother-infant interactions that prepare infants for language.

The present study provided insights into the ecology of an everyday activity, such as diapering, and the role that language and action play in this natural setting that is common in many cultures. The data analysis provided a detailed, systematic description and annotation of maternal multimodal responsive behavior in early natural mother-infant interactions. The behavioral classes elaborated after the fine-grained coding focused on the practices, which the mothers-infant dyads in the data shared in common. Although the variability was addressed in the analysis, it was not the main focus. Future analyses could concentrate on the individual differences between mother-infant dyads and make the individual developmental trajectories of the dyads visible. This would shed even more light on the ways in which development and the education of attention are shaped by the cumulative history of the interaction on the dyads (Hsu & Fogel, 2003b), as relating multimodal practices to different interaction and verbalization styles might provide further interactive factors influencing infant learning.

Also, the analysis could be expanded to include more resources used in the interactions. For the mothers, this could include the use of prosody and touch. For the infants, this could include the analysis of smiling or infant vocalizations. These were mentioned in the qualitative analyses of the interactional sequences, as they were made relevant in the specific examples. The next step would be to develop behavioral measures that would be coded throughout the entire corpus and would be used as a basis for further statistical analyses.

Finally, future studies could expand on the findings of this observational study by testing them experimentally. For example a study could address the multiple types of temporal overlap identified. There the suggestion was that the various forms of co-occurrence might actually provide

a differential structuring, or packaging, of actions or language. It would be interesting to find out whether different structuring influences what is being learned, by making different aspects of the stimuli salient. Also, concerning the switching roles-model and its predictive role for language acquisition and the acquisition of verbs, it would be interesting to experimentally control for the precise effects of the two variables — infant attention to synchrony and the mother's synchronic response to their infant's gaze — on language development at two years of infants' age.

References

Adamson, L. B. (1995). *Communication development during infancy.* Boulder, CO: Westview Press.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff & K. Hirsh-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). Oxford: Oxford University Press.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental psychology*, *36*(2), 190–201.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*(3), 99–102.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, *55*, 1278–1289.

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. New York, NY: Cambridge University Press.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of experimental child psychology*, *64*(1), 3–26.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). New York, NY: Psychology Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(04), 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Bateson, M. C. (1979). The epigenesis of conversational interaction: a personal account of research development. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 63–77). Cambridge, UK: Cambridge University Press.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an

innate gaze module? evidence from human neonates. *Infant Behavior and Development*, *23*(2), 223–229.

Bell, S. M., & Ainsworth, M. D. S. (1972). Infant crying and maternal responsiveness. *Child development*, *43*(4), 1171–1190.

Bloom, L. (1995). *The transition from infancy to language: Acquiring the power of expression*. New York: Cambridge University Press.

Boersma, P., & Weenik, D. (2010). Praat: doing phonetics by computer [version 5.1.30] [Computer program]. Retrieved from `http://www.praat.org/`

Bornstein, M. H., & Tamis-LeMonda, C. S. (1989). Maternal responsiveness and cognitive development in children. *New Directions for Child and Adolescent Development*, *1989*(43), 49–61.

Bornstein, M. H., & Tamis-Lemonda, C. S. (1997). Maternal responsiveness and infant mental abilities: Specific predictive relations. *Infant Behavior and Development*, *20*(3), 283–296.

Bornstein, M. H., Tamis-LeMonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., … Vardi, D. (1992). Maternal responsiveness to infants in three societies: The united states, france, and japan. *Child Development*, *63*(4), 808–821.

Bortfeld, H., Shaw, K., & Depowski, N. (2013). Disentangling the influence of salience and familiarity on infant word learning: methodological advances. *Frontiers in psychology*, *4*, 175.

Brand, R. J., & Tapscott, S. (2007). Acoustic packaging of action sequences by infants. *Infancy*, *11*(3), 321–332.

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (Vol. 71, pp. 29–55). Chichester, U.K: Wiley-Blackwell.

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

Bruner, J. S. (1983). *Child's talk: Learning to use language*. New York, N.Y: W W Norton &

Company Incorporated.

Bullowa, M. (1979). Prelinguistic communication: a field for scientific research. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 1–62). Cambridge, UK: Cambridge University Press.

Buresh, J. S., Woodward, A., & Brune, C. W. (2006). The roots of verbs in prelinguistic action knowledge. In *Action meets word: How children learn verbs* (pp. 208–227). New York, NY: Oxford University Press.

Call, J., & Carpenter, M. (2002). Three sources of information in social learning. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 211–228). Cambridge, MA: MIT Press.

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., … others (2010). Integration of action and language knowledge: A roadmap for developmental robotics. *IEEE Transactions on Autonomous Mental Development*, *2*(3), 167–195.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, *63*(4), i–174.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, *20*(12), 553–557.

Clark, A. (1999). An embodied cognitive science? *Trends in cognitive sciences*, *3*(9), 345–351.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.

Colas, A. (1999). Introducing infants to referential events: a development study of maternal ostensive marking in french. *Journal of child language*, *26*(1), 113–131.

Colombo, J., & Bundy, R. S. (1983). Infant response to auditory familiarity and novelty. *Infant Behavior and Development*, *6*(2), 305–311.

Condon, W. S., & Sander, L. W. (1974). Neonate movement is synchronized with adult speech:

Interactional participation and language acquisition. *Science*, *183*(4120), 99–101.

Cowley, S. J. (2011). Taking a language stance. *Ecological Psychology*, *23*(3), 185–209.

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

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.), *Processes of change in brain and cognitive development. attention and performance xxi* (pp. 249–274). Oxford: Oxford University Press.

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

de Barbaro, K., Johnson, C., Forster, D., & Deak, G. (2013). Methodological considerations for investigating themicro-dynamics of social interaction development. *IEEE Transactions on Autonomous Mental Development*, *5*(3), 358–370.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, *14*(10), 441–447.

Demuth, C., & Mey, G. (2015). Qualitative methodology in developmental psychology. In J. Wright (Ed.), *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier.

Dent, C. H. (1990). An ecological approach to language development: An alternative functionalism. *Developmental Psychobiology*, *23*(7), 679–703.

Edelman, G. M. (1987). *Neural darwinism: The theory of neuronal group selection.* New York, NY: Basic Books.

Farran, D. C., Hirschbiel, P., & Jay, S. (1980). Toward interactive synchrony: The gaze patterns of mothers and children in three age groups. *International Journal of Behavioral Development*, *3*(2), 215–224.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605.

Feldman, R. (2011). Maternal touch and the developing infant. In M. J. Hertenstein & S. J. Weiss (Eds.), *The handbook of touch: neuroscience, behavioral, and health perspectives* (pp. 373–407). New York, NY: Springer Publishing Company.

Ferguson, C. A. (2004). Talking to children: A search for universals. In B. Lust & C. Foley (Eds.), *First language acquisition: The essential readings* (Vol. 4, pp. 176–189). Oxford, UK: Blackwell Publishing.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3-and 4-month-old infants: an advantage of words over tones. *Child development*, *81*(2), 472–479.

Filipi, A. (2009). *Toddler and parent interaction: The organisation of gaze, pointing and vocalisation*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Filipi, A. (2013). Withholding and pursuit in the development of skills in interaction and language. *Interaction Studies*, *14*(2), 139–159.

Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, *27*(2), 181–194.

Fodor, J. A. (1975). *The language and thought*. Cambridge, MA: Harvard University Press.

Fogel, A. (1977). Temporal organization in mother-infant face-to-face interaction. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 119–152). New York, NY: Academic Press.

Fogel, A. (1988). Cyclicity and stability in mother-infant face-to-face interaction: A comment on cohn and tronick (1988). *Developmental Psychology*, *24*(3), 393–395.

Fogel, A. (1992). Co-regulation, perception and action: Reply to reaction. *Human movement science*, *11*(4), 505–523.

Fogel, A. (1993a). *Developing through relationships*. Chicago, IL: University of Chicago Press.

Fogel, A. (1993b). Two principles of communication: Co-regulation and framing. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 9–22).

London, UK: Routledge.

Fogel, A. (2006). Dynamic systems research on interindividual communication: The transformation of meaning-making. *Journal of developmental processes*, *1*(1), 7–30.

Fogel, A., & Garvey, A. (2007). Alive communication. *Infant Behavior and Development*, *30*(2), 251–257.

Fogel, A., Garvey, A., Hsu, H.-C., & West-Stroming, D. (2006). *Change processes in relationships: a relational-historical research approach*. New York, NY: Cambridge University Press.

Friedlander, B. Z. (1968). The effect of speaker identity, voice inflection, vocabulary, and message redundancy on infants' selection of vocal reinforcement. *Journal of Experimental Child Psychology*, *6*(3), 443–459.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, *22*(3-4), 455–479.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT: Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Routledge.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–922.

Gogate, L. J., & Bahrick, L. E. (2001). Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations. *Infancy*, *2*(2), 219–231.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child development*, *71*(4), 878–894.

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable–object pairings: The role of temporal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 508–519.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension: an ecological–dynamic systems view. *Developmental Science*, *4*(1), 1–18.

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces carving up events for learning language. *Perspectives on Psychological Science*, *5*(1), 33–42.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith from associationist to social sophisticate. *Current Directions in Psychological Science*, *15*(1), 30–33.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York, NY: Halsted Press.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of pragmatics*, *32*(10), 1489–1522.

Gredebäck, G., Melinder, A., & Daum, M. (2010). The development and neural basis of pointing comprehension. *Social neuroscience*, *5*(5-6), 441–450.

Grimm, H., Aktas, M., & Frevert, S. (2000). *Sprachentwicklungstest für zweijährige kinder (setk-2)*. Göttingen: Göttingen: Hogrefe.

Grimm, H., & Doil, H. (2000). *Elternfragebögen für die früherkennung von risikokindern: Elfra*. Göttingen: Hogrefe, Verlag für Psychologie.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509–516.

Grossmann, T., Parise, E., & Friederici, A. D. (2010). The detection of communicative signals directed at the self in infant prefrontal cortex. *Frontiers in human neuroscience*, *4*.

Harris, M., Jones, D., Brookes, S., & Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British Journal of Developmental Psychology*, *4*(3), 261–268.

Harris, M., Jones, D., & Grant, J. (1983). The nonverbal context of mothers' speech to infants. *First Language*, *4*(10), 21–30.

Heft, H. (1989). Affordances and the body: An intentional analysis of gibson's ecological approach to visual perception. *Journal for the theory of social behaviour*, *19*(1), 1–30.

Henning, A., Striano, T., & Lieven, E. V. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, *28*(4), 519–536.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: The MIT Press.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child development*, *73*(2), 418–433.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., … Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the society for research in child development*, *65*, i–135.

Holmlund, C. (1995). Development of turntakings as a sensorimotor process in the first 3 months: A sequential analysis. In K. Nelson & Z. Reger (Eds.), *Children's language* (Vol. 8, pp. 41–64). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hsu, H.-C., & Fogel, A. (2001). Infant vocal development in a dynamic mother-infant communication system. *Infancy*, *2*(1), 87–109.

Hsu, H.-C., & Fogel, A. (2003a). Social regulatory effects of infant nondistress vocalization on maternal behavior. *Developmental Psychology*, *39*(6), 976–991.

Hsu, H.-C., & Fogel, A. (2003b). Stability and transitions in mother-infant face-to-face commu-

nication during the first 6 months: a microhistorical approach. *Developmental psychology*, *39*(6), 1061–1082.

Hsu, H.-C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences□. *Infant behavior and development*, *24*(1), 107–128.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. the dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11-12), 11–12.

Jaffe, J., Stern, D. N., & Peery, J. C. (1973). "conversational" coupling of gaze behavior in prelinguistic human development. *Journal of Psycholinguistic Research*, *2*(4), 321–329.

Jean, A. D., Stack, D. M., & Fogel, A. (2009). A longitudinal investigation of maternal touching across the first 6 months of life: Age and context effects. *Infant Behavior and Development*, *32*(3), 344–349.

Kärtner, J., Keller, H., & Yovsi, R. D. (2010). Mother–infant interaction during the first 3 months: The emergence of culture-specific contingency patterns. *Child development*, *81*(2), 540–554.

Kaye, K. (1980). Why we don't talk 'baby talk' to babies. *Journal of Child Language*, *7*(03), 489–507.

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology*, *16*(5), 454–464.

Kaye, K., & Wells, A. J. (1980). Mothers' jiggling and the burst—pause pattern in neonatal feeding. *Infant Behavior and Development*, *3*, 29–46.

Keller, H. (2000). Sozial-emotionale grundlagen des spracherwerbs. In H. □. Grimm (Ed.), *Enzyklopädie der psychologie, themenbereich c, serie iii, band 3: Sprachentwicklung* (pp. 379–402). Göttingen [u.a.]: Hogrefe, Verl. für Psychologie.

Keller, H., & Gauda, G. (1987). Eye contact in the first months of life and its developmental

consequences. In R. H & S. H. C (Eds.), *Psychobiology and early development* (pp. 129–143). Amsterdam: Elsevier.

Keller, H., Lohaus, A., Völker, S., Cappenberg, M., & Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Development*, *70*(2), 474–485.

Koester, L. S., Papoušek, H., & Papoušek, M. (1989). Patterns of rhythmic stimulation by mothers with three-month-olds: A cross-modal comparison. *International Journal of Behavioral Development*, *12*(2), 143–154.

Kopp, S., & Bergmann, K. (2010). When sequential gestures construct a spatial discourse. In *4th conference of the international society for gesture studies (isgs)* (p. 54). Frankfurt/Oder, Germany.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York, NY: Basic books.

Lavelli, M., & Fogel, A. (2005). Developmental changes in the relationship between the infant's attention and emotion during early face-to-face communication: the 2-month transition. *Developmental psychology*, *41*(1), 265.

Legerstee, M. (2005). *Infants' sense of people: precursors to a theory of mind*. Cambridge, UK: Cambridge University Press.

Lewkowicz, D. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological bulletin*, *126*(2), 281.

Lewkowicz, D., & Kraebel, K. (2004). The value of multisensory redundancy in the development of intersensory perception. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 655–678). Cambridge, MA: MIT press.

Li, X. (2013). Language and the body in the construction of units in mandarin face-to-face interaction. In *Units of talk–units of action* (Vol. 25, pp. 343–375). Amsterdam, NL: John Benjamins Publishing.

Longhi, E. (2009). Songese': Maternal structuring of musical interaction with infants. *Psychol-*

ogy of Music, *37*(2), 195–213.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083.

Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*(2), 235–250.

Majid, A., Bowerman, M., Staden, M. v., & Boster, J. S. (2007). The semantic categories of cutting and breaking events: A crosslinguistic perspective. *Cognitive Linguistics*, *18*(2), 133–152.

Mandler, J. M. (1988). How to build a baby: On the development of an accessible representational system. *Cognitive Development*, *3*(2), 113–136.

Mandler, J. M. (1992). How to build a baby: Ii. conceptual primitives. *Psychological review*, *99*(4), 587.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*(1), 3–36.

Markova, G., & Legerstee, M. (2006). Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology*, *42*(1), 132–141.

Masataka, N. (2003). *The onset of language*. Cambridge, UK: Cambridge University Press.

McGillion, M., Herbert, J., Pine, J., Keren-Portnoy, T., Vihman, M., & Matthews, D. (2013). Supporting early vocabulary development: What sort of responsiveness matters? *IEEE Transactions on Autonomous Mental Development*, *5*(3), 240-248.

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

Messer, D., & Vietze, P. (1984). Timing and transitions in mother-infant gaze. *Infant Behavior and Development*, *7*(2), 167–181.

Meyer, M., Hard, B., Brand, R. J., McGarvey, M., & Baldwin, D. A. (2011). Acoustic packaging: Maternal speech and action synchrony. *IEEE Transactions on Autonomous Mental*

Development, *3*(2), 154–162.

Muir, D., & Field, J. (1979). Newborn infants orient to sounds. *Child development*, *50*, 431–436.

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

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274.

Nagai, Y., & Rohlfing, K. (2009). Computational analysis of motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 44–54.

Nebel, B., & Bürckert, H.-J. (1995). Reasoning about temporal relations: a maximal tractable subclass of allen's interval algebra. *Journal of the ACM (JACM)*, *42*(1), 43–66.

Nomikou, I., & Rohlfing, K. (2011). Language does something: Body action and language in maternal input to three-month-olds. *IEEE Transactions on Autonomous Mental Development*, *3*(2), 113–128.

Nomikou, I., Rohlfing, K., & Szufnarowska, J. (2013). Educating attention: Recruiting, maintaining, and framing eye contact in early natural mother-infant interactions. *Interaction Studies*, *14*(2), 240–267.

Papoušek, H., & Papoušek, M. (1987). Intuitive parenting: A dialectic counterpart to the infant's integrative competence. In J. D. E. Osofsky (Ed.), *Handbook of infant development* (pp. 669–720). Oxford, UK: John Wiley & Sons.

Papoušek, H., Papoušek, M., & Bornstein, M. (2002). Intuitive parenting. In M. H. Bornstein (Ed.), *Handbook of parenting volume 2 biology and ecology of parenting* (pp. 183–203). Mahwah, NJ: Laurence Erlbaum Associates, Publishers.

Papoušek, M. (1994). *Vom ersten schrei zum ersten wort: Anfänge der sprachentwicklung in der vorsprachlichen kommunikation*. Bern, Göttingen, Toronto, Seattle: Verlag Hans Huber.

Papoušek, M. (1995). Origins of reciprocity and mutuality in prelinguistic parent-infant 'dia-

logues.'. In I. Marková, C. F. Graumann, & K. Foppa (Eds.), *Mutualities in dialogue* (pp. 58–81). Cambridge, UK: Cambridge University Press Cambridge.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*(6), 137–157.

Papoušek, M., & Papoušek, H. (1990). Excessive infant crying and intuitive parental care: Buffering support and its failures in parent-infant interaction. *Early Child Development and Care*, *65*(1), 117–126.

Pecher, D., & Zwaan, R. A. (2005). The role of perception and action in memory, language, and thinking. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition.* New Your, NY: Cambridge University Press.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, *308*(5727), 1430–1430.

Pitsch, K., Vollmer, A.-L., & Muehlig, M. (2013). Robot feedback shapes the tutor's presentation how a robot's online gaze strategies lead to micro-adaptation of the human's conduct. *Interaction Studies*, *14*(2), 268–296.

Pitsch, K., Vollmer, A.-L., Rohlfing, K., Fritsch, J., & Wrede, B. (2014). Tutoring in adult-child-interaction: On the loop of the tutor's action modification and the recipient's gaze. *Interaction Studies*, 55–98.

Plunkett, K. (1997). Theories of early language acquisition. *Trends in cognitive sciences*, *1*(4), 146–153.

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681.

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child development*, *77*(2), 266–280.

Psathas, G. (1995). *Conversation analysis: The study of talk-in-interaction* (Vol. 35). Thousand Oaks, CA: Sage.

Rączaszek-Leonardi, J. (2009). Symbols as constraints: the structuring role of dynamics and self-organization in natural language. *Pragmatics & Cognition*, *17*(3), 653–676.

Rączaszek-Leonardi, J., Nomikou, I., & Rohlfing, K. (2013). Young children's dialogical actions: The beginnings of purposeful intersubjectivity. *IEEE Transactions on Autonomous Mental Development*, *5*(3), 210-221.

Rader, N. d. V., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, *10*(2-3), 202–221.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188–194.

Rochat, P. (2001). Dialogical nature of cognition. *Monographs of the Society for Research in Child Development*, *66*(2), 133–143.

Rogoff, B., Mistry, J., Göncü, A., Mosier, C., Chavajay, P., & Heath, S. B. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development*, *58*(236), i–179.

Rohlfing, K. (2013). *Frühkindliche semantik. eine einführung*. Tübingen: Narr Verlag.

Rohlfing, K., & Deak, G. (2013). Microdynamics of interaction: Capturing and modeling infants' social learning. *IEEE Transactions on Autonomous Mental Development*, *5*(3), 189-191.

Rohlfing, K., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science*, *15*(3), 426–435.

Rolf, M., Hanheide, M., & Rohlfing, K. (2009). Attention via synchrony: Making use of multimodal cues in social learning. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 55–67.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*(2), 221–234.

Sachs, J. (1977). The adaptive significance of linguistic input to prelinguistic infants. In C. Snow & C. A. Ferguson (Eds.), *Talking to children. language input and acquisition. cambridge:*

Cup (pp. 51–61). New York, NY: Cambridge University Press.

Schaffer, H. R. (1984). *The child's entry into a social world.* London, UK: Academic Press.

Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis* (Vol. 1). New York, NY: Cambridge University Press.

Schillingmann, L., Wrede, B., & Rohlfing, K. (2009). A computational model of acoustic packaging. *IEEE Transactions on Autonomous Mental Development*, *1*(4), 226–237.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., … Uhmann, S. (2009). Gesprächsanalytisches transkriptionssystem 2 (gat 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, *10*, 353–402.

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

Smith, L. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review*, *25*(3), 278–298.

Smith, L., Colunga, E., & Yoshida, H. (2010). Knowledge as process: contextually cued attention and early word learning. *Cognitive science*, *34*(7), 1287–1314.

Smith, L., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial life*, *11*(1-2), 13–29.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Snow, C. (1986). Conversations with children. In P. Fletcher & M. Garman (Eds.), *Language acquisition : studies in first language development* (pp. 69–89). Cambridge, UK: Cambridge University Press.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of child language*, *4*(1), 1–22.

Stack, D. M., & Jean, A. D. (2011). Communicating through touch: Touching during parent–infant interactions. In M. J. Hertenstein & S. J. Weiss (Eds.), *The handbook of touch: neuroscience, behavioral, and health perspectives* (pp. 273–298). New York, NY: Springer

Publishing Company.

Stack, D. M., LePage, D. E., Hains, S., & Muir, D. W. (1996). Qualitatitive changes in maternal touch as a function of instructional condition during face-to-face social interactions. *Infant Behavior and Development*, *19*, 761.

Stack, D. M., & Muir, D. W. (1992). Adult tactile stimulation during face-to-face interactions modulates five-month-olds' affect and attention. *Child Development*, *63*(6), 1509–1525.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses.* Cambridge, MA: The MIT Press.

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & R. L. A (Eds.), *The effect of the infant on its caregiver.* Oxford, UK: Wiley-Interscience.

Stern, D. N. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York, NY: Basic Books.

Stern, D. N. (2000). Putting time back into our considerations of infant experience: a microdiachronic view. *Infant mental health journal*, *21*(1-2), 21–28.

Stern, D. N., Beebe, B., Jaffe, J., & Bennett, S. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 177–202). New York, NY: Academic Press.

Stern, D. N., & Gibbon, J. (1979). Temporal expectations of social behavior in mother-infant play. In E. B. Thoman & E. Hillsdale (Eds.), *Origins of the infant's social responsiveness* (pp. 409–429). NJ: Erlbaum.

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in cognitive sciences*, *10*(10), 471–476.

Suliteanu, G. (1979). The role of songs for children in the formation of musical perception. In J. Blacking & J. W. Kealiinohomoko (Eds.), *The performing arts: music and dance* (Vol. 10, pp. 205–219). The Hague: Mouton.

Sullivan, J. W., & Horowitz, F. D. (1983). Infant intermodal perception and maternal multimodal stimulation: Implications for language development. In C. Rovee-Collier & L. L. P (Eds.), *Advances in infancy research* (pp. 183–239). Norwood: Ablex.

Takada, A. (2011). Pre-verbal infant-caregiver interaction. In A. Duranti, E. Ochs, & B. Schieffelin (Eds.), *The handbook of language socialization* (Vol. 71, pp. 56–80). Chichester, UK: John Wiley & Sons.

Tamis LeMonda, C., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, *72*(3), 748–767.

Thelen, E. (2000). Grounded in the world: Developmental origins of the embodied mind. *Infancy*, *1*(1), 3–28.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thorngate, W. (1986). The production, detection, and explanation of behavioral patterns. In J. Valsiner (Ed.), *The individual subject and scientific psychology* (pp. 71–93). New York, NY: Plenum Press.

Toda, S., Fogel, A., & Kawai, M. (1990). Maternal speech to three-month-old infants in the united states and japan. *Journal of Child Language*, *17*(02), 279–294.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–690.

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

Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? from individual to shared to collective intentionality. *Mind and Language*, *18*(2), 121–147.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, *4*(12), 197–211.

Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 43–78). Norwood, NJ: Ablex Publishing Corporation.

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

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* (pp. 121–173). New York, NY: Cambridge University Press.

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of child psychology and psychiatry*, *42*(1), 3–48.

Tulbert, E., & Goodwin, M. H. (2011). Choreographies of attention: Multimodality in a routine family activity. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (Vol. 8, pp. 79–92). New York, N.Y: Cambridge University Press.

Turing, A. (2004). Intelligent machinery (1948). In B. J. Copeland (Ed.), *The essential turing* (pp. 395–432). New York, NY: Oxford University Press.

Turkewitz, G., & McGuire, I. (1978). Intersensory functioning during early development. *International Journal of Mental Health*, *7*(1/2), 165–182.

Vallotton, C. D. (2009). Do infants influence their quality of care? infants' communicative gestures predict caregivers' responsiveness. *Infant Behavior and Development*, *32*(4), 351–365.

Valsiner, J. (2000). Data as representations: Contextualizing qualitative and quantitative research strategies. *Social science information*, *39*(1), 99–113.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech

for young infants. *Developmental science*, *7*(3), 270–276.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes.* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Watson, J. (1985). Contingency perception in early social development. In T. Field & F. NA (Eds.), *Social perception in infants* (pp. 157–176). Norwood, NJ: Ablex.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*(4), 625–636.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: a professional framework for multimodality research. In *Proceedings of lrec* (Vol. 2006).

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.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving*. *Journal of child psychology and psychiatry*, *17*(2), 89–100.

Woodward, A. L. (2004). Infants' use of action knowledge to get a grasp on words. In (pp. 149–172). Cambridge, MA: MIT Press.

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, *34*(1), 72–90.

Yoder, P. J., Warren, S. F., McCathren, R., & Leew, S. (1998). Does adult responsivity to child behavior facilitate communication development? In A. M. Wetherby, S. F. Warren, & J. Reichle (Eds.), *Transitions in prelinguistic communication* (Vol. 7, pp. 39–58). Dexter, MI: Paul H. Brookes Pub co.

Zukow-Goldring, P. (1990). Socio-perceptual bases for the emergence of language: An alternative to innatist approaches. *Developmental psychobiology*, *23*(7), 705–726.

Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early development and parenting*, *5*(4), 195–211.

Zukow-Goldring, P. (1997). A social ecological realist approach to the emergence of the

lexicon: Educating attention to amodal invariants in gesture and speech. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development: Ecological approaches to organism–environment systems.* (pp. 199–252). Washington, DC: American Psychological Association.

Zukow-Goldring, P. (2006). Assisted imitation: Affordances, effectivities, and the mirror system in early language development. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 469–500). New York, N.Y: Cambridge University Press.

Zukow-Goldring, P., & Rader, N. (2005). Discussion to the symposium 'multimodal motherese'. In *10th international congress for the study of child language.* Berlin, Germany.

A. Appendix for Chapter 4

Descriptive Statistics for GAZE

This section includes the complete results presented in Chapter 4. It includes the data for the interactions with the three- and six-month-old infants.

Table A.1.: Descriptive Statistics for GAZE at three months

Variable		N	Mean	Std. Dev.	Minimum	Maximum
M-gaze at 3M	(%)	17	52.91	16.33	24.93	75.61
	(Avg. in sec.)	17	3.62	2.03	1.10	8.31
I-gaze at 3M	(%)	17	58.29	20.90	21.25	93.95
	(Avg.in sec.)	17	10.26	8.05	2.95	33.77
M-gaze at object 3M	(%)	17	40.99	17.14	14.30	74.00
	(Avg. in sec.)	17	2.71	1.37	1.23	6.28
I-gaze at object 3M	(%)	17	18.35	19.10	0.00	58.24
	(Avg. in sec.)	17	4.13	4.15	0.00	13.59
M-gaze away 3M	(%)	17	2.16	2.35	0.00	7.70
	(Avg. in sec.)	17	1.35	0.94	0.00	3.70
I-gaze away 3M	(%)	17	12.89	11.74	1.08	47.00
	(Avg. in sec.)	17	3.44	3.03	0.54	12.93

Table A.2.: Descriptive Statistics GAZE at six months

Variable		N	Mean	Std. Dev.	Minimum	Maximum
M-gaze at 6M	(%)	17	46.96	20.26	5.31	87.40
	(Avg. in sec.)	17	2.62	2.24	0.87	10.57
I-gaze at 6M	(%)	17	32.89	16.66	5.87	69.25
	(Avg. in sec.)	17	3.61	1.65	1.21	7.80
M-gaze at object 6M	(%)	17	45.71	21.33	8.38	84.10
	(Avg. in sec)	17	2.62	1.83	0.91	9.01
I-gaze at object 6M	(%)	17	31.00	17.78	4.54	60.09
	(Avg. in sec.)	17	4.24	3.73	1.67	17.66
M-gaze away 6M	(%)	17	1.91	1.66	0.00	5.80
	(Avg. in sec.)	17	1.35	0.92	0.00	3.70
I-gaze away 6M	(%)	17	22.66	10.00	10.70	46.58
	(Avg. in sec.)	17	3.62	1.69	1.61	7.80

Table A.3.: Descriptive Statistics EYE CONTACT at three months

Variable		N	Mean	Std. Dev.	Minimum	Maximum
eye contact 3M	(%)	17	34.84	14.03	8.61	66.05
	(Avg. in sec.)	17	3.01	1.79	0.97	7.08

Table A.4.: Descriptive Statistics EYE CONTACT at six months

Variable		N	Mean	Std. Dev.	Minimum	Maximum
eye contact 6M	(%)	17	20.53	13.66	1.69	43.38
	(Avg. in sec.)	17	1.80	0.76	0.73	3.13

GAZE pairwise comparisons for three- and six-month-old infants

This section presents the complete data of the comparison of gaze patterns between the two age groups. Marked in color are the significant differences between the two age groups.

Table A.5.: Paired Sample *t* - test for GAZE over interaction time at three and six months

Variable Pairs	*t*	*df*	Sig. (2-tailed)
eye contact 3M and 6M	4.152	16	**.001**
M-gaze at 3M and 6M	1.557	16	.139
M-gaze at obj. 3M and 6M	-1.263	16	.225
M-gaze away 3M and 6M	.440	16	.666
I-gaze at 3M and 6M	4.118	16	**.001**
I-gaze away 3M and 6M	-2.319	16	**.034**
I-gaze at obj. 3M and 6M	-2.432	16	**.027**

Table A.6.: Paired Sample *t* - test for average duration of GAZE events at three and six months

Variable Pairs	*t*	*df*	Sig. (2-tailed)
eye contact 3M and 6M	3.814	16	**.002**
M-gaze at 3M and 6M	2.947	16	**.009**
M-gaze at obj. 3M and 6M	.199	16	.845
M-gaze away 3M and 6M	.006	16	.995
I-gaze at 3M and 6M	3.414	16	**.004**
I-gaze away 3M and 6M	-.223	16	.827
I-gaze at obj. 3M and 6M	-.074	16	.942

B. Appendix for Chapter 6

Synchrony over interaction time

This section includes the descriptive statistics of the coded behavioral classes over the entire interaction time. It includes the complete data for each modality and for the overall synchrony.

Table B.1.: Results for synchrony over interaction time at three months

Modality	N	Mean	*SD*	Minimum	Maximum
hands	17	16.97	7.53	5.24	31.55
head	17	7.99	5.01	1.08	18.21
face	17	5.68	4.07	0.00	13.11
torso	17	3.90	2.24	0.56	9.16
lips	17	.97	1.25	0.00	4.18

Table B.2.: Results for average duration of synchrony events at three months

Modality	N	Mean	*SD*	Minimum	Maximum
hand	17	1.47	0.36	0.82	2.26
face	17	1.13	0.43	0.00	1.92
head	17	1.05	0.33	0.63	2.04
torso	17	1.13	0.28	0.54	1.60
lips	17	0.86	0.88	0.00	3.36

Table B.3.: Results for overall synchrony over interaction time at three months

Modality	N	Mean	*SD*	Minimum	Maximum
overall synchrony	17	29.1	10.1	13.93	43.87

Table B.4.: Results for synchrony over interaction time at six months

Modality	N	Mean	*SD*	Minimum	Maximum
face	17	10.84	8.32	0.00	31.67
hands	17	15.54	7.89	0.00	28.18
head	17	5.18	4.90	0.00	20.39
lips	17	0.91	0.83	0.00	2.57
torso	17	9.27	5.84	0.00	20.24

Table B.5.: Results for average duration of synchrony events at six months

Modality	N	Mean	*SD*	Minimum	Maximum
face	17	0.87	0.29	0.00	1.49
hands	17	0.94	0.29	0.00	1.39
head	17	0.52	0.23	0.00	1.06
lips	17	0.82	0.96	0.00	4.36
torso	17	0.73	0.25	0.00	1.14

Table B.6.: Results for overall synchrony over interaction time at six months

Modality	N	Mean	*SD*	Minimum	Maximum
overall synchrony	17	27.06	12.15	0.00	44.25

SYNCHRONY pairwise comparisons for three- and six-month-old infants

This section includes the complete data for the pairwise comparison of the proportion of the interaction time in which mothers used intermodal synchrony. It also includes the complete data for the pairwise comparison of the durations of synchrony events for the three- and six-month-old infants. Marked n color are the significant comparisons.

Table B.7.: Paired Sample *t* - test for synchrony over interaction time at three and six months

Variable Pairs	*t*	*df*	Sig. (2-tailed)
face 3M and 6M	-3.54	16	**.003**
hands 3M and 6M	1.054	16	.307
head 3M and 6M	3.177	16	**.006**
lips 3M and 6M	.199	16	.845
torso 3M and 6M	-4.764	16	**.000**
overall 3M and 6M	1.018	16	.324

Table B.8.: Paired Sample *t* - test for average duration of synchrony events at three and six months

Variable Pairs	*t*	*df*	Sig. (2-tailed)
face 3M and 6M	2.263	16	**.038**
hands 3M and 6M	7.323	16	**.000**
head 3M and 6M	6.692	16	**.000**
lips 3M and 6M	.142	16	.889
torso 3M and 6M	5.385	16	**.000**

Synchrony over spoken time

This section includes the results of the synchrony in relation to maternal vocal behavior.

Table B.9.: Results for synchrony over spoken time at three months

Modality	N	Mean	*SD*	Minimum	Maximum
face	17	14.08	9.05	0.00	35.02
hands	17	42.80	13.92	18.31	62.99
head	17	19.09	9.43	4.60	35.84
lips	17	2.60	3.65	0.00	13.49
torso	17	10.30	6.18	1.77	25.45

Table B.10.: Results for overall synchrony over spoken time at 3M

Modality	N	Mean	*SD*	Minimum	Maximum
overall synchrony	17	73.34	12.71	54.20	94.69

Table B.11.: Results for synchrony over spoken time at six months

Modality	N	Mean	*SD*	Minimum	Maximum
face	17	27.23	17.50	0.00	62.42
hands	17	40.46	17.87	0.00	67.44
head	17	12.37	9.32	0.00	40.19
lips	17	2.44	2.15	0.00	6.26
torso	17	24.28	14.73	0.00	61.86

Table B.12.: Results for overall synchrony over spoken time at 6M

Modality	N	Mean	*SD*	Minimum	Maximum
overall synchrony	17	69.73	22.92	0.00	93.84

Pairwise comparisons for synchrony over spoken time for the three- and six-month-old infants

This section includes the complete data for the pairwise comparison of the proportion maternal behavior which was reinforced by body movement. Marked in color are the significant comparisons.

Table B.13.: Paired Sample *t* - test for synchrony over spoken time at three and six months

Variables Pairs	*t*	*df*	Sig. (2-tailed)
face 3M and 6M	-4.085	16	**.001**
hands 3M and 6M	.603	16	.555
head 3M and 6M	2.922	16	**.010**
lips 3M and 6M	.167	16	.870
torso 3M and 6M	-5.309	16	**.000**
overall 3M and 6M	.598	16	.558

Overlap of multiple modalities

This section includes the co-occurrence of multiple modalities in the data. It presents all possible combinations of modalities, and the percentages of durations over spoken time. Results are presented for sets of two, three, and four modalities.

Table B.14.: Co-occurring modalities in combinations of two at three months

Modality combination	N	Mean	*SD*	Minimum	Maximum
Hands-Face	17	2.87	3.50	0.00	13.63
Hands-Torso	17	4.01	3.36	0.39	12.89
Hands-Head	17	3.64	3.54	0.00	11.62
Hands-Lips	17	0.12	0.32	0.00	1.31
Face-Torso	17	2.11	2.29	0.00	9.22
Face-Head	17	7.26	4.89	0.00	15.75
Face-Lips	17	0.02	0.04	0.00	0.12
Torso-Head	17	2.37	2.28	0.00	9.57
Torso-Lips	17	0.22	0.44	0.00	1.49
Head-Lips	17	0.05	0.16	0.00	0.45

Vocal types

This section includes the frequency of occurrence of vocal types in all modalities.

In the following, the cross-tabulation data are included for age and the vocal types. Marked in color are variables with the biggest difference between observed and expected frequencies.

Relationship between synchrony and vocal type classes

This sections includes the cross-tabulation data for synchrony modalities and vocal types. Marked in color are variables with the biggest difference between observed and expected frequencies.

Table B.15.: Co-occurring modalities in combinations of two at six months

Modality combination	N	Mean	*SD*	Minimum	Maximum
Hands-Face	17	9.55	7.40	0.00	23.19
Hands-Torso	17	9.89	5.97	0.00	24.90
Hands-Head	17	3.01	2.61	0.00	10.50
Hands-Lips	17	0.38	0.55	0.00	2.12
Face-Torso	17	7.94	7.15	0.00	25.75
Face-Head	17	7.30	7.22	0.00	31.03
Face-Lips	17	0.33	0.57	0.00	1.70
Torso-Head	17	4.33	5.78	0.00	24.80
Torso-Lips	17	0.48	0.57	0.00	1.98
Head-Lips	17	0.35	0.47	0.00	1.18

Table B.16.: Co-occurring modalities in combinations of three at three months

Modality combination	N	Mean	*SD*	Minimum	Maximum
Hands-Face-Torso	17	0.80	1.33	0.00	5.53
Hands-Face-Head	17	0.99	1.24	0.00	3.83
Hands-Head-Lips	17	0.01	0.05	0.00	0.19
Face-Torso-Head	17	1.03	1.13	0.00	3.87
Hands-Torso-Head	17	0.68	1.13	0.00	4.45
Torso-Head-Lips	17	0.00	0.00	0.00	0.00
Face-Hands-Lips	17	0.00	0.00	0.00	0.00
Lips-Head-Face	17	0.00	0.00	0.00	0.00
Hands-Torso-Lips	17	0.00	0.00	0.00	0.00
Face-Torso-Lips	17	0.00	0.00	0.00	0.00

Table B.17.: Co-occurring modalities in combinations of three at six months

Modality combination	N	Mean	SD	Minimum	Maximum
Hand-Face-Torso	17	3.77	3.55	0.00	11.48
Hands-Face-Head	17	2.02	2.41	0.00	8.65
Hands-Head-Lips	17	0.01	0.03	0.00	0.13
Face-Torso-Head	17	2.28	2.62	0.00	9.96
Hands-Torso-Head	17	1.57	1.75	0.00	6.43
Torso-Head-Lips	17	0.09	0.16	0.00	0.60
Face-Hands-Lips	17	0.05	0.19	0.00	0.80
Lips-Head-Face	17	0.09	0.17	0.00	0.50
Hands-Torso-Lips	17	0.17	0.55	0.00	2.27
Face-Torso-Lips	17	0.18	0.41	0.00	1.30

Table B.18.: Co-occurring modalities in combinations of four at three months

Modality combination	N	Mean	SD	Minimum	Maximum
Hands-Face-Torso-Head	17	0.26	0.38	0.00	0.96
Hands-Face-Torso-Lips	17	0.00	0.00	0.00	0.00
Face-Torso-Head-Lips	17	0.00	0.00	0.00	0.00
Head-Lips-Face-Hands	17	0.00	0.00	0.00	0.00
Hands-Torso-Head-Lips	17	0.00	0.00	0.00	0.00

Table B.19.: Co-occurring modalities in combinations of four at six months

Modality combination	N	Mean	*SD*	Minimum	Maximum
Hands-Face-Torso-Head	17	0.79	0.92	0.00	2.88
Hands-Face-Torso-Lips	17	0.05	0.19	0.00	0.77
Face-Torso-Head-Lips	17	0.03	0.08	0.00	0.30
Head-Lips-Face-Hands	17	0.001	0.01	0.00	0.02
Hands-Torso-Head-Lips	17	0.01	0.01	0.00	0.05

Table B.20.: Frequency of occurrence of vocal types at three months

Vocal Type	N	Mean	*SD*	Minimum	Maximum
speech	17	72.1	13.93	48.73	98.18
emotionalization	17	12.61	7.83	0	24.90
sound	17	8.67	6.75	0	20.90
imitation	17	5.01	3.95	0	11.17
song	17	1.62	4.38	0	17.26

Table B.21.: Frequency of occurrence of vocal types at six months

Vocal Type	N	Mean	*SD*	Minimum	Maximum
speech	17	61.79	20.58	0	82.69
emotionalization	17	12.38	6.30	0	31.05
sound	17	12.36	8.26	0	33.33
imitation	17	5.69	5.50	0	17.95
song	17	1.9	6.29	0	26.15

Table B.22.: Crosstabulation analysis AGE vs. VOCAL TYPE

			emot.	imit.	song	sound	speech	Total
			Vocal Types					
Age	3M	Count	21	9	3	**14**	**99**	146
		Expected Count	21.3	9.9	2.8	**17.4**	**94.6**	146.0
	6M	Count	39	19	5	**35**	**168**	266
		Expected Count	38.7	18.1	5.2	**31.6**	**172.4**	266.0
Total		Count	60	28	8	49	267	412
		Expected Count	60.0	28.0	8.0	49.0	267.0	412.0

Table B.23.: Cross-tabulation analysis for body class vs. vocal type at three months

			face	hands	head	lips	torso	Total
					Body Classes			
Vocal Types	speech	Count	15	**45**	28	0	10	98
		Expected Count	18.4	**37.4**	28.6	2.7	10.9	98.0
	emotionalization	Count	**6**	6	6	0	2	20
		Expected Count	**3.8**	7.6	5.8	0.6	2.2	20.0
	sound	Count	2	3	3	**4**	2	14
		Expected Count	2.6	5.3	4.1	**0.4**	1.6	14.0
	imitation	Count	**3**	0	**4**	0	1	8
		Expected Count	**1.5**	3.1	**2.3**	0.2	0.9	8
	song	Count	1	1	1	0	1	4
		Expected Count	0.8	1.5	1.2	0.1	0.4	4.0
Total		Count	27	55	42	4	16	144
		Expected Count	27.0	55.0	42.0	4.0	16.0	144.0

Table B.24.: Crosstabulation analysis for body class vs. vocal type at six months

| | | | Body Classes | | | | | |
			face	hands	head	lips	torso	Total
Vocal Types	speech	Count	35	**68**	30	0	35	168
		Expected Count	40.1	**58.5**	29.9	4.5	35.0	168.0
	emotionalization	Count	**13**	13	5	0	7	38
		Expected Count	**9.1**	13.2	6.8	1.0	7.9	38.0
	sound	Count	6	7	6	**7**	9	35
		Expected Count	8.4	12.2	6.2	**0.9**	7.3	35.0
	imitation	Count	**8**	3	**5**	0	3	19
		Expected Count	**4.5**	6.6	**3.4**	0.5	4	19
	song	Count	1	1	1	0	1	4
		Expected Count	1.0	1.4	0.7	0.1	0.8	4.0
Total		Count	63	92	47	7	55	264
		Expected Count	63.0	92.0	47.0	7.0	55.0	264.0

C. Appendix for Chapter 7

This section includes the paired comparisons for the overlap of the synchrony modalities with the infants' gaze at the mother and the infants' gaze away. The significant differences and significant correlations are marked in color.

Table C.1.: Paired Sample t - test for synchrony overlap with infant gaze

Variables Pairs	t	df	Sig. (2-tailed)
face infant gaze at & infant gaze away	11.31	16	**<.001**
hands infant gaze at & infant gaze away	3.77	16	**.002**
head infant gaze at & infant gaze away	11.54	16	**<.001**
lips infant gaze at & infant gaze away	2.95	16	**.009**
torso infant gaze at & infant gaze away	5.76	16	**<.000**

Table C.2.: Correlations of the average durations of synchrony-infant gaze overlap with synchrony

Face		Face-gaze at	Face-gaze away
	Pearson *r*	**.962**	.473
	Sig. (2-tailed)	**.000**	.055
	N	17	17
Hands		Hands-gaze at	Hands-gaze away
	Pearson *r*	**.802**	**.721**
	Sig. (2-tailed)	**>.000**	**.001**
	N	17	17
Head		Head-gaze at	Head-gaze away
	Pearson *r*	**.976**	.283
	Sig. (2-tailed)	**>.001**	.271
	N	17	17
Torso		Torso-gaze at	Torso-gaze away
	Pearson *r*	**.757**	**.611**
	Sig. (2-tailed)	**>.001**	**.009**
	N	17	17
Lips		Lips-gaze at	Lips-gaze away
	Pearson *r*	**.856**	**.246**
	Sig. (2-tailed)	**>.001**	.359
	N	17	17

D. Appendix for Chapter 8

Table D.1.: Individual measures for three-month-olds

Subj. No.	Maternal synchrony	Infants' attention to synchrony	Attended synchrony
01	43.88	56.84	54.45
02	21.60	43.03	61.56
03	13.93	13.06	19.92
04	30.05	33.03	83.13
05	43.64	47.95	52.30
06	42.37	44.52	63.64
07	27.41	29.97	71.63
08	21.67	26.39	69.84
09	22.53	25.17	97.51
10	43.74	56.55	43.77
11	30.76	31.08	94.94
12	33.02	40.31	77.94
13	20.35	19.89	77.98
14	36.50	41.77	73.21
MIN	13.93	13.06	19.92
MAX	43.87	56.83	97.51
MEAN (SD)	30.82 (10.1)	36.40 (13.1)	67.27 (20.5)

Table D.2.: Individual measures for six-month-olds

Subj. No.	Maternal synchrony	Infants' attention to synchrony	Attended synchrony
01	44.25	59.49	75.51
02	26.04	32.63	34
03	17.94	33.55	24.12
04	32.94	37.26	48.09
05	41.93	53.98	58.28
06	44.02	58.28	38.33
07	11.66	28.74	29.33
08	15.59	28.51	46.49
09	20.31	29.83	23.47
10	34.44	43.56	59.29
11	30.66	32.98	33.97
12	24.08	29.74	46.47
13	0	0	0
14	30.22	41.48	45.70
MIN	0	0	0
MAX	44.25	59.49	75.51
MEAN (SD)	26.72 (12.9)	36.43 (15.1)	40.22 (18.5)

Table D.3.: Individual measures ELFRA at 24M
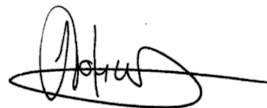
Subj. No.	ELFRA-2 prod.	ELFRA-2 verbs%	ELFRA-2 nouns %
01	200	14.50	50.50
02	174	14.37	59.77
03	66	4.55	74.24
04	45	6.67	77.78
05	191	16.75	52.36
06	160	14.38	60.00
07	99	6.06	79.80
08	215	18.14	50.23
09	27	0	85.19
10	81	8.64	76.54
11	136	11.03	68.38
12	210	18.57	51.43
13	16	6.25	68.75
14	157	17.20	57.32
MIN	16	0	50.23
MAX	215	18.57	85.19
MEAN (SD)	126.93 (70.17)	11.22 (5.87)	65.16 (12.14)

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig ohne fremde Hilfe verfasst habe und keine anderen Quellen oder Hilfsmittel als die angegebenen verwendet habe. Wörtliche oder sinngemäß übernommene Zitate sind als solche gekennzeichnet. Datensätze, Zeichnungen Skizzen und graphische Darstellungen sind selbständig erstellt.

Portsmouth, 22.05.2018

Ort, Datum   Unterschrift