Vocal Interactions at the Dawn of Communication: The Emergence of Mutuality and Complementarity in Mother-Infant Interaction*

Giuseppe Leonardi, Iris Nomikou, Katharina J. Rohlfing and Joanna Rączaszek-Leonardi

Abstract— **Continuous interaction of mother and infant in the first weeks and months of an infant's life entrains the infant on many crucial aspects of how to do things together. Contingencies of gaze, vocalizations, and other movements are slowly routinized; this scaffolds directing of attention to each other and the world and gives to such multimodal interactions meaning. It is within these continuous interactions with caregivers that language emerges, starting from the first non-reflexive vocalizations that infants produce. The response that caregivers promptly give to these vocalizations informs infants of their relevance and helps shape them. We explored this systematicity by observing the coupling of infants' and mothers' vocalizations in unconstrained interactions longitudinally. While at three months, mothers seem to answer consistently to any speech related vocalization within the first two seconds, this pattern fades away at six and eight months. What remains stable across age is a structure in which overlapping vocalizations are rare and give way to a sequential pattern of vocal reciprocity — an embryonic turn-taking behavior. Discussion relates this finding to early coordination in other modalities in an attempt to sketch a more holistic account of emerging co-action.**

I. INTRODUCTION

On the way to language acquisition, for a child, a "fair amount of early learning" [1, p. 68] is about the mastery of joint action. Within the thousands of interactions experienced day after day, infants learn how to communicate effectively to achieve joint goals. Bruner [2] suggests that out of social responses to each other, a reciprocity develops, together with a pattern of interaction that the infant anticipates. Such repeated routinized interactions construct a constrained, systematic, predictive setting within which the infant can learn to co-act with others [3]. This is true for multiple resources used in interaction: for the coordination of attention, the development of the ability to use gestures and verbal behavior consistently as a conventionalized system, and the use of language as a form of social practice.

For the development of vocal behavior, the progression to language develops from initially neutral sounds in terms of communicative value. Yet, these occur in a rich social environment, which selects and interprets many of them as communicative acts [1, 4]. Caregivers answer to infants' vocalizations, and these temporal contingencies teach infants that their acts are meaningful [3, 5, 6, 7, 8]. It is within these interactions that sounds are slowly shaped into the phonological forms of infants' linguistic environment and acquire functional value as generators of different kinds of responses from the caregivers and as actions with which joint goals can be achieved [2, 9].

The essence of our approach lies in viewing interaction not as a mechanism of sending and receiving messages, but as a dynamical process of generating and transforming meaning together [10, 11]. In this view, interaction is not a mere context in which (any) development takes place, but part of the process of development itself. Accordingly, development takes place as interacting parties adapt to each other and coordinate their behavior, forming a qualitatively new, systemic level of analysis that is not solely based on imitation [8]. Both caregivers and infants contribute to the experience of joint interaction; they do not act as individuals but as a system. At the heart of such a systemic view is the notion that the elements composing the system (in this case caregiver and infant), apart from standing in some relation to one another, affect each other by their participation in the system [12]. To investigate development means to take into account how this mutual influence might happen [13].

This view suggests that coordination within interaction is key for development [14, 15, 16]. The way in which it may cause developmental change is by viewing it as operating on different time-scales: the first time-scale involves the moment-to-moment adjustments that keep an interaction going; on the second time-scale, these dynamics act on multiple repeated interactions within a cumulative history of experiences [14, 17]. On this time scale, certain patterns of coordination are stabilized, routinized and become shared social practices [2].

In existing approaches, coordination usually means the *sameness* between the behavior of mother and infant. Mutuality allows for mother and infant to 'share' subjective states and emotional expressions [18] as mother and infant are attuned [19]. Rączaszek-Leonardi and colleagues 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 [3, p. 211].

Another way of looking at coordination is as a *dialogue*. Here, great value is put on the allocation of different roles in interaction for the achievement of one overarching goal for a dyad within a given situation [11, 20]. In this view, the complementarity of mothers' and infants' behaviors becomes central [14, 17].

In both modes of coordination in development (imitative or complementary), relative timing of participants' behaviors,

G. Leonardi is with the Adam Mickiewicz University, Poznań 61-712 Poland (phone: +48-61-829-2215; e-mail: giuleonardi@ gmail.com).

I. Nomikou, is with Paderborn University, Paderborn 33098 Germany (e-mail: iris.nomikou@uni-paderborn.de).

K. J. Rohlfing is with Paderborn University, Paderborn 33098 Germany (e-mail: katharina.rohlfing@uni-paderborn.de).

J. Rączaszek-Leonardi is with the University of Warsaw, Warsaw 00-187 Poland (e-mail: raczasze@psych.uw.edu.pl).

i.e. the way one's behavior unfolds as a function of the time-dependent phases in another's behavior, appears to be crucial. Much research has already turned on the *when* of interaction. Infants' sensitivity to timing has been investigated, revealing that early in development infants can both influence the timing of caregivers' behavior towards them as well as be influenced by it. Jaffe and collaborators [15] showed how already 4-month-old infants time the pauses of their vocalizations. Research has further shown that timing extends beyond vocal behavior to include kinesic modalities [21]. Crucially, infants are not only receivers of temporal patterns but also actively contribute to their emergence [22].

Recently, work inspired by the dynamical, systemic views on development presented above has turned to the use of continuous measurements and the extraction of dynamical measures (as in Cross-Recurrence Quantification Analysis, CRQA [23]). This work has shed light on the dynamics of caregiver-child interactions [see e.g. 24, 25], providing new insights not only into coordination processes but also into the emergence of specific interaction patterns such as the identification of who is leading in the interaction. In [25] we used this approach on longitudinal data, uncovering the emergence of coordination within the first year of an infant's life. In this study on the development of gaze-at-face coordination, we found that there seems to be no leader or follower of the timing of gaze, pointing to the fact that mutual attention is a co-regulated process [17]. Furthermore, we found that within development, although the amount of time spent in mutual gaze decreases in absolute terms, the gaze coordination becomes more efficient.

Yet, it remains unclear how interaction patterns are stabilized in other modalities, and whether coordination is modality-dependent. Gaze seems to be an unquestionably powerful channel for informational coupling with the world and others [25; but see 26]. However, especially in the social world, it is by no means the only modality. Not less importantly, coordination is reported also in vocal behavior. Specifically, Jaffe and colleagues [15] have shown that 4-month-old infants are already capable of timing their vocal behavior to that of their caregivers. Moreover, in a recent study on turn-taking Gratier et al. [22] showed that 2–4 month-old infants engage in vocal exchanges with minimal switching pauses, suggesting that they are already capable of projecting the end of their mother's turn to be able to begin theirs after a minimal pause. This finding shows that very young infants already have an active role in vocal coordination. By looking at the development of turn-taking in infants', Hilbrink et al. [27] found that in the first five months infants' overlapping vocalizations remain quite stable, but from the fifth month on the overlap starts to decrease, giving way to a more alternate form of vocal coordination. They also observed that early in development infants were relatively fast to respond to mothers' vocalizations, but around the age of nine months they become slower. These studies are interesting in themselves, but the next step would be to consider the dynamic timing of vocal production in the dyad, which could tell us whether, e.g. there is a leader or a follower in the coordination of vocal behavior and how these roles develop with age.

With the present paper, we focus on early vocal interactions between mothers and infants in the first year of an infant's life. Our first aim is to explore how the first protophones [28], i.e. the non-reflexive vocal behavior in infants are treated by their caregivers. More precisely, we investigate whether mothers promote spontaneous speech-related sound production in infants by systematically vocalizing in reply to them. Our second aim is to examine whether there is a temporal coupling between mothers' and infants' vocalizations and sketch its development over the first eight months of infants' age. Our final aim is to relate the coordination patterns in the vocal modality with our findings on the coordination of gaze of the same mother-infant dyads.

Based on previous work [e.g. 24], we hypothesize that mothers will promote spontaneous speech-related sound production in infants by systematically vocalizing in reply to them. It is reasonable to assume that infant vocalizations will receive a modality-specific response in order to educate the child towards the use of language. When considering the turn-taking patterns which are ubiquitous across human cultures, two things become obvious: the tight timing structure (most of the turns are made within around 200-300ms) [29], and the fact that there is hardly any overlap in speech in a dialogue. We expected to see developmental change as the child's sensitivity to possibilities in vocal behavior emerges.

Our research goal can also contributes to the design of systems that will be able to learn from progressive dialogical coordination. Robotic developments have benefitted from the concept of scaffolding [e.g. 30] as a form of tutoring behavior. Scaffolding is assumed to reduce the complexity of the learning information in the input [31]. Below, we aim to provide concrete insights into how infants become scaffolded via modality-specific responses into the dynamics of dialogical coordination, which is affected by the level of learner's competence/development.

II. METHOD

A. Participants

For the present analysis we used the data corpus described in [32]. This includes longitudinal naturalistic observations of 17 German mother-infant dyads filmed during an everyday routine activity, i.e., diaper changing. Data collection was made at six monthly visits to the families' homes starting when the infants were three months old and until they became eight months old. The present analysis includes the interactions of 12 mother-infant dyads at three time points: at three (3;12), six (6;4) and eight (8;0) months of age.

B. Coding

Data were coded manually using Praat phonetics transcription software [33], to mark the onset and offset of mothers' and infants' vocalizations. For the mothers we coded any verbal behavior. For the infants we followed the approach presented in [28] and coded only protophones, i.e. infant pre-speech vocalizations other than reflexive sounds such as cries, laughs and vegetative sounds. As in earlier research [21], a vocalization was defined as a continuous vocal behavior bounded by pauses of at least 0.3 sec.

TABLE I. AVERAGE GAPS AND OVERLAPS OF VOCALIZATION TURNS (IN MILLISECONDS, STANDARD DEVIATION IN BRACKETS)

Visit	Mothers		Infants	
	Gaps	*Overlaps*	*Gaps*	*Overlaps*
Visit 1	564.3 (582.6)	-569.4 (909.8)	755.4 (692.7)	-767.7 (872.4)
Visit 2	527.6 (594.9)	-473.7 (491.4)	708.8 (701.7)	-682.5 (587.2)
Visit 3	613.4 (676.3)	-602.5 (630.7)	810.7 (764.0)	-688.3 (514.4)

C. Data and Analyses

Turn-taking analysis involved the extraction of the gaps and overlaps across mothers' and infants' vocalizations from the Praat coded output files. The gaps were computed from the offset of a vocalization to the onset of the following vocalization of the other dyad member, and adopting the criteria used in [22] we considered only gaps less than 3000ms long. In case vocalizations overlapped, the overlapping period was computed in the same way as the gaps, i.e. from offset to onset of dyad vocalizations, resulting hence in this case in a negative value. The average gaps and overlaps for mothers and infants at the three visits are presented in Table 1.

Turn-taking data is an aggregate measure which compactly illustrates the average delays across infants' and mothers' vocal exchanges and corroborates previous reports of studies on mother-infant turn-taking behavior [22, 27]. Yet these average delays do not capture the full dynamical interweaving of vocalizations and pauses of the dyad. In order to do that, we separately constructed the time series of mothers' and infants' vocalizations, in binary form. Accordingly, periods of vocal production were marked with '1' and pauses/silence were marked with '0'. We sampled this behavior with a frequency of 2 Hz (a time resolution of 500ms). On these time series we computed CRQA using the *crqa* package in R [34], and extracted the diagonal-wise recurrence rate as a function of lag [34, 35] from each interaction of every dyad (one for every visit, at three, six and eight months of age).

CRQA matches like events in two related time series at every point in time of the series [23, 35]. Positive matches constitute a recurrence. The amount of recurrence obtained indexes the level of coordination of mothers and infants, i.e. the coordination of the vocalizations. From the analysis it is possible to extract the so called diagonal-wise recurrence profile, i.e. the recurrence as a function of the lag between events in the time series. When the events mapped in the time series consistently follow each other at a certain lag, we would observe an increased amount of recurrence at that lag, otherwise no particular structure would emerge. Fig. 1 shows the diagonal-wise recurrence profile of the averaged recurrence rate of the 12 dyads analyzed, when the infants were three, six and eight months old. The central point at lag 0 represents simultaneous behaviors, where recurrence would indicate the same behavior at the same time; on the right side at positive lags we have the recurrence of events where the infant followed the vocalization of the mother with a vocalization, while on the left side, at negative lags, it is the mother's vocalization which follows the vocalization of the infant.
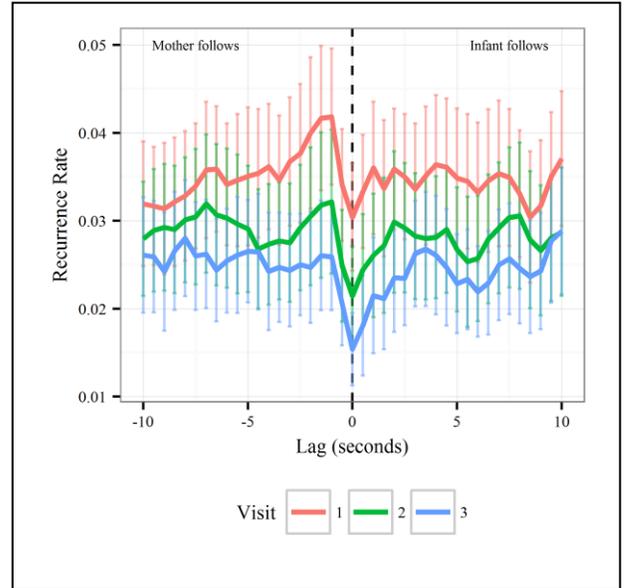

Figure 1. Averaged diagonal-wise recurrence profiles for the vocalization interactions of infants and caregivers at Visit 1 (3-months-old infants), Visit 2 (6-months-old infants) and Visit 3 (8-months-old infants). Vertical lines indicate the standard error.*)*

The analysis of these profiles was then framed in terms of linear mixed models, separately run on the negative lags (mother follows - left side of Fig. 1) and the positive lags (infant follows - right side of Fig. 1). In this way, we can separately evaluate if the slopes in the two halves of the profile are significantly different from zero and if they change across visits. As fixed effects we considered the Lag and the Visit while Dyad was entered as the only random effect. Dyad was allowed to vary both at the intercept and at the slopes of the fitted fixed effects. The negative lags in the recurrence profiles were changed in sign to allow a straightforward comparison between the estimated coefficients for the right-side and left-side profiles analyses (i.e. positive coefficients would signify increasing recurrence for longer lags in both cases).

III. RESULTS

A. Turn-taking data

The gap and overlap times data were analyzed with a two-factor split-plot ANOVA with the between-subjects factor being Dyad Member (with two levels: mother or infant) and the within-subjects factor being Visit (with three levels). In both cases the only significant effect was the main effect of Dyad Member (gap times: $F(1,33)=35.71$, $p<0.001$; overlap times: $F(1,33)=7.55$, $p<0.01$). Mothers' vocalizations followed infants' vocalizations at a consistently shorter gap than vice versa, and overlaps were also consistently shorter for mothers than for infants. This effect did not change from the first to the third visit, and the interaction between Visit and Dyad Member did not reach significance in both analyses

B. CRQA data

The averaged recurrence profiles show two patterns which need to be analyzed: a valley in the profile at lag 0 and a hump in recurrence on the left side of the profile. We analyzed the first one by means of a two-way within subjects

ANOVA. Within subjects factors were Visit (3 levels) and Lag (2 levels: lag 0 vs. other lags). The only significant effect was the main effect of Lag ($F_{(1, 11)}=12.03$, $p<0.01$) while neither the effect of Visit or Visit per Lag interaction reached significance. This supports the idea that the level of recurrence at lag 0 is indeed significantly lower than the other lags in average.

The linear mixed models analyses were aimed at checking the second pattern in the profiles (the presence of the hump). For the right-side section (infant follows) of the recurrence profiles, the only significant coefficient is the Intercept ($b=0.041$, $t(12)=3.879$, $p<0.01$), while neither Lag or Visit or their interaction produce any significant result. This means that while the amount of recurrence is significantly different from zero, the slopes for changing levels of lag or for the visits at different ages are rather flat, and do not significantly differ between them, in spite of the observed tendency towards decreasing levels of recurrence at every visit. On the other side, for the left-side section (mother follows) of the profiles, we register significant (i.e. different from zero) coefficients for the Intercept ($b=0.046$, $t(12)=4.12$, $p<0.01$), for the Lag ($b=-0.0012$, $t(152)=-3.051$, $p<0.01$) and for the interaction of Lag and Visit ($b=0.0005$, $t(683)=3.029$, $p<0.01$). In other words, the amount of recurrence is again significantly different from zero, and as Lag increases, generally recurrence decreases but do so differently for the different levels of Visit: at visit 2 and 3, the lag slope tends to be flatter and the recurrence hump in the first few seconds from an infant vocalization disappears.

IV. DISCUSSION

In our study, we explore early mother-infant vocal interactions in infants' first year of life. Our aim is to investigate how these first non-reflexive infants' vocalizations are treated by their caregivers. More precisely, we hypothesize that mothers promote spontaneous speech-related sound production in infants by systematically reacting to them in a variety of ways. We reason that these responses might be part of a process in which children's behavior is shaped towards particular conventions by means of regularities in vocalization timing (among other behaviors) in interactions [36].

To address our aims and hypotheses, we analyze the dynamics of mother-infant vocalizations recorded during a natural everyday routine activity, such as diapering, when the infants were three, six and eight months old. From the recurrence profiles, we find that already in early interactions, turn-taking is taking place. This is evidenced by the minimum in the recurrence profile corresponding to the lag 0, indicating that overlapping vocalizations are the least occurring temporal pattern in the data. As we can see in Figure 1, the average percentage of recurrence at lag 0 seems to decrease from visit to visit. The decrease is visible in the 'valley' in the profile, which becomes more pronounced with age, in comparison to the rest of the profile, although not significantly so. Statistically significant instead, is the difference between recurrence at lag 0 and averaged recurrence at other lags. At lag 0, minimum recurrence of vocalizations is registered and then in average either both infant and mother are not vocalizing or one is vocalizing and the other waits. This finding echoes the findings of Hilbrink

and colleagues [27], who documented a decrease in overlapping turns from 5 months onward. This coordination of listening and vocalizing may be due to the emergence of the first signs of what will shape the complementary roles of talker and listener in a series of turn-taking behaviors in a conversation. Clearly, this reasoning is only based on the observation of the averaged profiles of diagonal-wise recurrence, and future work will aim to a more punctual analytical treatment of it.

From the profile, we also derive the observation that recurrence seems to peak within the first two seconds of an infant's vocalization: Mothers tend to respond to infant vocalizations, and they do it promptly as has been already reported in the literature [24]. Yet our findings suggest that this behavior is most reliable in the first visit at three months of age [see also 37] than in the following visits, as it was additionally corroborated by the linear mixed model results. This finding suggests that at three months, mothers are clearly taking up a follower role within a very specific temporal window, since the recurrence profile shows a peak at this lag range. A possible explanation is that mothers actively scaffold the production of infants' vocalizations providing contingent responses to the non-reflexive sounds produced by the infants in an attempt to elicit further 'talk'. Such an intuitive strategy was employed in an experimental setup and found to be beneficial for the shaping of infants' speech-related vocal development at babbling age [5, 38]. As evident in the profile, the fact that this peak gradually disappears from visit to visit suggests that as the infants become older, mothers tend to respond to their vocalizations with looser coordination and with more variable timing. On the other hand, we should account for a possible alternative interpretation of this results. Due to the way the analysis was performed, we cannot completely rule out that the missing humps in the recurrence profile on the side of 'mothers following' in visit 2 and 3, or on the side of 'infants following' in all the visits, could have been determined by the increasing variability of vocalization lengths of infants and the mothers alike. The observation that mothers' behavior becomes more variable with infants' development was already made by Fogel [39]. Mothers might be actively trying to establish vocal exchange as a reliable communicative resource early in development. Later though, their behavior might become more flexible, allowing for more variability in responding. Longer and/or more variable vocalization times could then produce shifts of the recurrence registered in the single instances of the lag dependent profiles and hence, possibly, could also mask any systematic structuring in the profiles in average (i.e. the humps). We will address this possible masking effect in future analyses aimed at normalizing vocalizations lengths of the interacting dyad [see e.g. 24].

Yet another possibility could be that the gradually disappearing peak in the mothers' vocalization reactions is related to the quality of infant vocalizations. Existing studies have shown that mothers intuitively react in different ways to different phonological properties of infant vocalizations [7]. It could well be that with three months old infants, mothers promptly respond to any kind of infant vocalization, whereas with older infants, they may be responding readily to more advanced infant vocalizations (such as consonant vowel

clusters, or reduplicated syllables) to encourage and scaffold them [2], while reducing their responsiveness to more primitive vocalizations (such as quasi-vowels). We are currently pursuing this analysis.

When looking to the right side of the profile (positive lags), infants do not seem to show any time-locked response to mothers' vocalizations. In other words, infants do follow with vocalizations to their mothers', as evidenced by the amount of recurrence in the profile (the significant Intercept in the linear mixed model), but on average, their response times are more variable. This might explain why we do not see any peak of recurrence at a particular lag on the right-side of the recurrence profile, which again calls for a preliminary normalization of vocalization times in CRQA [24]. Our finding echoes both Gratier et al. [22] and Hilbrink et al. [27] results confirming infants' capacity of contingent vocalizations to mothers' vocalizations. However, our analysis complement these findings by showing that there is no specific or regular timing in these contingencies. The flatter profile indicates that an infant reply may occur with similar probability at any lag, and the coordination on the part of the infant seems to be very unspecific at all ages.

It is also interesting to note that, while in turn-taking data an average gap interval between vocalizations can always be obtained, a dynamically oriented analysis and the diagonal-wise recurrence profiles tell us a slightly different story. If coordination is indeed present with similar timing in all or most of the dyads then a definite structure will emerge, otherwise recurrence will spread evenly across lags pointing maybe to more idiosyncratic time contingencies across dyads. So aggregate statistics like the mean gaps and overlaps may be unable to detect real (i.e. universal) interactional structures as those probed by the recurrence profile. Only with the latter method, we notice that the prompt answer given by caregivers in the first visit (average gap 574.3ms), which in our traditional statistical analysis was not different from responses in the second and third visit, may have indeed unique interactional qualities and emerges as a clear hump in the recurrence profile.

An additional point worth discussing is the discrepancy between the value of the gaps from infant vocalization offset to mother vocalization onset (in average around 570ms across visits) and the peak in the recurrence profile, which seems to be shifted to later lags (around 1000-1500ms). We believe that much of this discrepancy is related to the time resolution of the coded time series, which, as said, sampled vocalization behavior every 500ms, and it could then be more apparent than real.

The recurrence analysis method not only revealed contingencies within interaction but it also evidenced specific interaction patterns of the dyad as a whole. This is the case with the recurrence level at lag 0. For some behaviors, such as gaze-at-face [25], we notice a progressive increase of the extent to which the infant's gaze matches the mother's, i.e. an increasing peak at lag 0 from visit to visit. Yet, in vocalizations, as expected, and as the present analysis showed, a different pattern emerges. The peak is clearly absent at lag 0, giving way to what seems like the progressive deepening of a 'valley' indexing avoidance to engage in the same behavior at the same time. Even though this pattern is interesting on its own, we do not think that a modular analysis of different modalities tells the whole story: the next fascinating question is therefore how the modalities relate to each other, becoming a pattern in time involving multiple different acts on both sides leading to a routine. This will be the next step in our research.

For the developments of intelligent learning systems, our results suggest that the ability to anticipate a number of sequentially organized actions is likely to be the basis for any approach to learning from interaction [40]. Our results are in accordance with key aspects, such as contingency, suggested by Cangelosi and colleagues [30] for models of social interaction. Furthermore, our results point to the fact that scaffolding is a dynamic process being strongly linked to the interaction history that the learner shares with the tutor. This is an aspect which is increasingly gaining attention [20]. This history allows a progression in so far as other behaviors might be responded to at the beginning of learning in contrast to later phases, in which the input will build on some competencies and target specific levels of knowledge. It is likely that the underlying representations of interaction change within this progression [40].

Moreover, in the context of the computational modeling literature, it could be interesting to confront computational models specifically built to capture the development of turn-taking behavior with the pattern of results presented here, where one of the vocalization agents (the mother) has already developed a vocalization overlapping avoidance action policy. Specifically models like the one proposed by Moulin-Frier and colleagues [41; see also 42] showing the self-organization of turn-taking behavior in agents just learning such behavior, would gain important empirical support if they also showed the contingent responsiveness within 2 seconds of infants' vocalizations, which emerged in our data particularly in the early phases of learning. Our data on the emergence of coordinative behavior in vocalizations (but see also [25]) can then serve as a crucial benchmark for this and other modeling strategies.

In sum, our analysis provided partial support to our hypothesis suggesting that mothers scaffold their infants towards the use of vocal means in their responses by answering promptly to them. This was only the case for the first data point, when the infants were three months old. For later visits, when the infants were six and eight months old, we found a more variable behavior of the mothers which needs to be disentangled by more targeted analyses. This pattern contrasts our recent research showing a rather tight coupling of mothers' and their infants' eye-gaze behaviors, which as an element of interaction [43] is related to later language development but becomes more efficient with infants' development [25]. We think that as infants entrain in a coordinative behavior, for example by responding more promptly to the mother's gaze directed towards them [25], they are also learning about interaction in general, which manifests itself in coupling some modalities while avoiding overlaps in others. We see a necessity to investigate multimodal responses at the dawn of communication as it seems likely that the modalities relate to and structure each other [32] conveying the experience of acting together.

REFERENCES

[1] J. S. Bruner, "Learning how to do things with words," in *Human growth and development*, J. Bruner, and A. Garton, Eds., 1978, pp. 62-84.

[2] J. S. Bruner, *Child's talk: Learning to use language*. New York, NY: W.W. Norton & Co. Inc, 1983.

[3] J. Rączaszek-Leonardi, I. Nomikou, and K. J. Rohlfing, "Young children's dialogical actions: The beginnings of purposeful intersubjectivity." *IEEE Trans. Auton. Ment. Dev.*, vol. 5(3), pp. 210-221, 2013.

[4] D. J. Lewkowicz, "Early experience and multisensory perceptual narrowing." *Dev. Psychob.*, vol. 56(2), pp.292-315, February 2014

[5] M. H. Goldstein, A. P. Kin, and M. J. West, "Social interaction shapes babbling: testing parallels between birdsong and speech," *Proc. Nat. Academy of Sciences of the USA*, vol. 100, pp. 8030-8035, 2003.

[6] M. H. Goldstein, J. A. Schwade, and M. H. Bornstein, "The value of vocalizing: five-month-old infants associate their own non-cry vocalizations with response from caregivers," *Child Development*, vol. 80, pp. 636-644, 2009

[7] J. Gros-Luis, M. J. West, M. H. Goldstein, and A. P. King, "Mothers provide differential feedback to infants' pre-linguistic sounds*," Int. J. of Behavioral Development*, vol. 30, pp.509-516, 2006.

[8] M. Asada, "Modeling early vocal development through infant-caregiver interaction: a review.*" IEEE Trans. Cog. Dev. Systems*, vol. 8(2), pp. 128–138, June 2016

[9] A. Lock, "The emergence of language," in *Action, gesture and symbol. The emergence of language*, A. Lock, Ed., 1978, pp.3-20.

[10] H. De Jaegher, and E. Di Paolo, "Participatory sense-making: An enactive approach to social cognition." *Phenomenology and the Cognitive Sciences*, vol. 6(4), pp. 485–507, December 2007.

[11] R. Fusaroli, J. Rączaszek-Leonardi, and K. Tylén, "Dialog as interpersonal synergy," *New Ideas Psych.*, vol. 32, pp. 146-157, 2014.

[12] M. Wilson, "Six views of embodied cognition." *Psychonomic Bulletin & Review*, vol. 9(4), pp. 625–636, December 2002.

[13] A. Fogel, *Developing through relationships*. Chicago, IL: University of Chicago Press. 1993.

[14] I. Nomikou, *The collaborative construction of early multimodal input and its significance for language development*, Unpublished Doctoral Thesis. Bielefeld University, 2015.

[15] J. Jaffe, B. Beebe, S. Feldstein, C. L. Crown, M. D. Jasnow, P. Rochat and D. N. Stern, "Rhythms of dialogue in infancy: Coordinated timing in development." *Monographs of the Society for Research in Child Development*, 2001.

[16] T. Fuchs, and H. De Jaegher, "Enactive intersubjectivity: Participatory sense-making and mutual incorporation." *Phenomenology and the Cognitive Sciences*, vol. 8(4), pp. 465-486, 2009.

[17] H-C. Hsu and A. Fogel, "Stability and transitions in mother-infant face-to-face communication during the first 6 months: a microhistorical approach." *Dev. Psych.*, vol. 39(6), pp. 1061-1082, 2003

[18] C. Trevarthen, "Communication and cooperation in early infancy: a description of primary intersubjectivity," in *Before speech: the beginning of interpersonal communication*, M. Bullowa, Ed., Cambridge, UK: Cambridge University Press, 1979, pp. 321-347.

[19] D. N. Stern, *The interpersonal world of the human infant*, New York, NY: Basic Books, 1985

[20] K. J. Rohlfing, B. Wrede, A.-L. Vollmer and P.-Y. Oudeyer, "An alternative to mapping a word onto a concept in language acquisition: pragmatic frames," *Frontiers in Cognitive Science*, vol. 7:470, 2016.

[21] D. N. Stern, B. Beebe, J. Jaffe, and S. Bennett, "The infant's stimulus world during social interaction: a study of caregiver behaviors with particular reference to repetition and timing," in *Studies in mother-infant interaction*, H. R. Schaffer, Ed., 1977, pp. 177-202.

[22] M. Gratier, E. Devouche, B. Guellai, R. Infanti, E. Yilmaz, and E. Parlato-Oliveira, "Early development of turn-taking in vocal interaction between mothers and infants," *Front. Psychol.*, 6:1167, 2015.

[23] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems." *Physics Reports*, vol. 438 (5-6), pp. 237-329, January 2007.

[24] A. S. Warlaumont, J. S. Richards, J. Gilkerson, and D. K. Oller, "A social feedback loop for speech development and its reduction in autism," *Psychological Science*, vol. 25, pp. 1314-1324, May 2014.

[25] I. Nomikou, G. Leonardi, K. J. Rohlfing and J. Rączaszek-Leonardi, "Constructing interaction: the development of gaze dynamics," *Infant and Child Development*, vol. 25(3), pp. 277-295, May 2016.

[26] K. de Barbaro, C.M. Johnson, D. Forster, and G.O. Deák, "Sensorimotor decoupling contributes to triadic attention: A longitudinal investigation of mother-infant-object interactions," *Child Development*, 2015.

[27] E. E. Hilbrink, M. Gattis, and S. C. Levinson, "Early developmental changes in the timing of turn-taking: a longitudinal study of mother-infant interaction," *Front. Psychol.*, 6:1492, 2015

[28] E. H. Buder, A. S. Warlamount, D. K. Oller, B. Peter, and A. A. N. MacLeod, "An acoustic catalog of prespeech vocalizations from a developmental perspective," *Comprehensive perspectives on child speech development and disorders: Pathways from linguistic theory to clinical practice*, pp. 103-134, 2013.

[29] M. Wilson, and T. P. Wilson, "An oscillatory model of the timing of turn taking," *Psychon. Bull. and Rev.*, vol. 12(6), pp. 957-968, 2005.

[30] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, et al., "Integration of action and language knowledge: A roadmap for developmental robotics." *IEEE Trans. on Auton. Ment. Dev.*, vol. 2(3), pp. 167–195, September 2010.

[31] K. J. Rohlfing, J. Fritsch, B. Wrede, and T. Jungmann, "How can multimodal cues from child-directed interaction reduce learning complexity in robots?" *Advanced Robotics*, vol. 20(10), pp. 1183–1199, 2006.

[32] I. Nomikou and K. J. Rohlfing, "Language does something: Body action and language in maternal input to three-month-olds." *IEEE Trans. on Auton. Ment. Dev.*, vol. 3(2), pp. 113-128, 2011.

[33] P. Boersma, and D. Weenik, "Praat: doing phonetics by computer [ver. 5.1.30]," [Computer program], 2010.

[34] M. Coco, and R. Dale, "Cross-recurrence quantification analysis of categorical and continuous time series: an R package." *Frontiers in Psychology*. 5:510. doi: 10.3389/fpsyg.2014.00510, 2014.

[35] R. Dale, A. S. Warlaumont, and D. C. Richardson, "Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams," *Int. J. Bifurcation and Chaos*, vol. 21, pp. 1153-1161, 2011

[36] J. Lany, and J. R. Saffran, "Statistical learning mechanisms in infancy." In *Comprehensive Developmental Neuroscience: neural circuit development and function in the brain, Vol. 3*, J.L.R. Rubenstein, and P. Rakic, Eds., Amsterdam: Elsevier, 2013, pp.231-248.

[37] H. Keller, A. Lohaus, S. Völker, M. Cappenberg and C. Athanasios, "Temporal contingency as an independent component of parenting behavior," *Child Development*, vol. 70, pp. 474-485, 1999.

[38] M. H. Goldstein, and J. A. Schwade, "Social feedback to infants' babbling facilitates rapid phonological learning," *Psychological Science*, vol. 19, pp. 515-523, 2008

[39] A. Fogel, "Temporal organization in mother-infant face-to-face interaction," *Studies in mother-infant interaction*, pp. 119-152, 1977.

[40] I. Nomikou, M. Schilling, V. Heller, and K. J. Rohlfing, "Language at all times. Action and interaction as contexts for enriching representations." *Interaction Studies*, vol. 17(1), pp. 120–145, 2016.

[41] C. Moulin-Frier, M. Sanchez-Fibla, and P. F. M. J. Verschure, "Autonomous development of turn taking behaviors in agent populations: a computational study." in *Proceedings of the 5th International Conference on Development and Learning and on Epigenetic Robotics*, Providence, RI, Aug 13-16, 2015, pp. 188-195.

[42] D. Y. Takahashi, D. Narayanan, and A. A. Ghazanfar, "A computational model for vocal exchange dynamics and their development in mormoset monkeys." *Philosophical Transactions of the Royal Society B*, vol. 371 (1963), p. 20150370

[43] K. J. Rohlfing and I. Nomikou, "Intermodal synchrony as a form of maternal responsiveness: association with language development," *Language, Interaction and Acquisition*, vol. 5(1), pp. 117-136, 2014.