On the loop of action modification and the recipient's gaze in adult-child interaction

Karola Pitsch, Anna-Lisa Vollmer, Jannik Fritsch, Britta Wrede, Katharina Rohlfing, Gerhard Sagerer

2009

Preprint:

This is an accepted article published in Gesture and Speech in Interaction. The final authenticated version is available online at: https://doi.org/[DOI not available]

On the loop of action modification and the recipient's gaze in adult-child-interaction

Karola Pitsch[1,2], Anna-Lisa Vollmer[2,1], Jannik Fritsch[2], Britta Wrede[1,2], Katharina Rohlfing[1,2], Gerhard Sagerer[1]

Bielefeld University, [1]Applied Informatics & [2]CoR-Lab
Bielefeld University, Faculty of Technology, P.O. Box 100 131, 33501 Bielefeld, Germany

karola.pitsch@uni-bielefeld.de; avollmer@cor-lab.uni-bielefeld.de

Abstract

Learning is a social endeavor, in which the learner generally receives support from his/her social partner(s). For instance, research in developmental psychology has demonstrated that - when talking to their children - tutors/adults not only modify their speech, but also their gestures and motions. However, analysis so far has focussed on the adult's presentation and was barely concerned with the recipient's (i.e. the child's) conduct. Also, the variability in parental behavior found less fine-grained analysis. In contrast, in this paper, we assume an interactional perspective investigating the loop between the tutor's and the learner's actions. With this approach, we aim both at discovering the levels and features of variability and at achieving a better understanding of how they come about within the course of the interaction. For our analysis, we used a combination of (1) qualitative investigation derived from ethnomethodological Conversation Analysis (CA), (2) semi-automatic computational 2D hand tracking and (3) a MATLAB based visualization of the data. Our analysis reveals that tutors not only shape their demonstrations differently with regard to the intended recipient per se (adult-directed vs. child-directed), but most importantly that the learner's feedback during the presentation is consequential for the concrete ways in which the presentation is carried out.

1 Introduction

Learning is a social and interactional endeavor, in which the learner generally receives support from his/her social environment. For instance, research in developmental psychology has demonstrated that – when talking or presenting new actions to their young infants – tutors/adults not only modify their speech, but also their gestures and motions (Brand et al. 2002). It has been suggested that these modifications scaffold children's acquisition of language and actions (Gogate et al. 2000, Brand et al. 2002). Recent studies have begun to identify objective criteria for gestural modification parameters using a computational solution (Rohlfing et al 2006, Vollmer et al 2009). They have been able to show that parents make longer pauses between different actions, perform their motion at a slower pace and decompose a rounded action trajectory into several linear movements when presenting a task, such as stacking differently sized cups, to their infants (age 8 to 11 months) as opposed to adult recipients. Other studies have attempted to determine whether such behaviour might indeed be consequential for the infant's learning and have undertaken experiments which suggest that infants prefer to look at CDI rather than at ADI action presentations (Brand et al 2008). While these studies provide a basis for systematically describing hand motion modifications (both for manipulative actions and gestures), one central aspect has been disregarded so far (cf. (Zukow-Goldring 1996): the *effect* of the modifications in the concrete interaction between the

adult/tutor and the infant/learner. How does the child – i.e. the recipient of the presented actions – respond to them? And how does the child's response feedback to the adult's action demonstration? We take an interactional perspective meaning that a child's (and everybody's) cognitive capabilties manifest themselves in the concrete and detailed ways of his/her conduct within an interaction. This conduct is visible for the adult/tutor, and the ways in which the child responds to some action presentation might, in turn, be consequential for the performance of the action itself. To perform some action within an interactional context is thus, essentially, a co-production of all participants involved. Recent research in Conversation Analysis reveals the effects which a co-participant's "online analysis" (Mondada 2006, Pitsch 2006) of an ongoing action has with regard to the action being performed: participants step-by-step adjust their own actions with regard to the visible (and audible) conduct of their co-participants. Against this background, we aim at investigating the *sources* of the variability in presenters' hand trajectories in the interaction between the participants.

2 Data and Method

2.1 Data

We use videotaped data from a semi-experimental setting, in which parents were asked to present a set of 10 manipulative tasks both to their infant and to another adult. During the tasks, a parent and the child were facing each other, sitting across a table. The situation was videotaped with two cameras (Rohlfing 2006). For the analysis presented here, we focus on a set of this data consisting of 12 parents and 8 children aged 8 to 11 months and one specific task: stacking differently sized cups, in which the assumed action consists of sequentially picking up the green, the yellow, and the red cup and to place them in the blue cup (cf. Fig. 1). The action of stacking cups represents a simple, manual task which requires the recipient to attentively observe the different parts of the action, i.e. to look at the right place at the right moment in time. Given the cognitive capabilities of young infants, an *explicit* task for the presenter consists in helping to orient the child's attention.

2.2 Method

To analyse the data, we used a combination of (1) qualitative investigation derived from ethnomethodological Conversation Analysis (CA), (2) semi-automatic computational 2D hand tracking and (3) a MATLAB based visualization of the data obtained. Using CA as the basic theoretical and methodological framework suggested an interactional perspective considering the adults' presentation of the task as a joint achievement of both, presenter and recipient. Consequently, we investigated the sequential organisation of the presenter's and recipient's actions (considering talk, manual actions/gestures, gaze orientation) and how they step-by-step react upon each other in the unfolding course of the interaction. In an initial sequential analysis of a small collection of cases we derived relevant analytical issues and categories "from the data themselves". On this basis, we decided to systematically capture the demonstrators' hand motions based on 2D computational pattern recognition methods (Vollmer et al 2009) and to transcribe/annotate the presenters' verbal actions and gaze direction as well as the recipient's gaze direction (ELAN). We used MATLAB for visualization of the hand trajectories and for linking them with the annotations.

3 Starting point: Variability of hand trajectories

If we examine the hand trajectories of the different groups on the basis of the plotted trajectories, we are able not only to find *generic* differences between the AAI-/ACI- conditions (e.g. more/less roundness {Rohlfing 2006},{Vollmer 2009}), but we can also identify a preliminary set of different trajectory *patterns*: (I) cases, in which the presenters' hand trajectories are flat without particularly marked points (Fig. 1a, 1c); (II) cases, in which the trajectories are a more pronounced and show a small peak towards the end (Fig. 1b); (III) cases, in which the presenters' hands perform some kind of peak or modulation at the onset (Fig. 1d, 1e); (IV) combinations of these trajectory types, especially those in which the first two stacking actions (green and yellow) show a high/pronounced shape; the third action (red line) is performed in a flat manner (Fig. 1e, 1f).

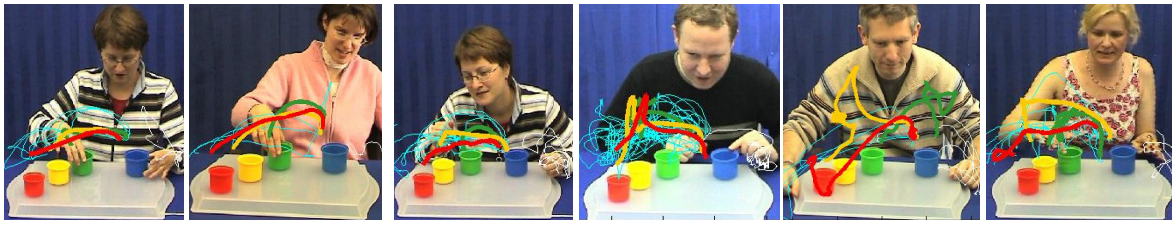


Fig. 1: Individual hand trajectories. Adult-Adult-Interaction (a, b) and Adult-Child-Interaction (c, d, e, f). Green/yellow/red trajectories mark the actions of stacking the cup of the corresponding colour into the blue one; other thin lines represent hand movements without cup.

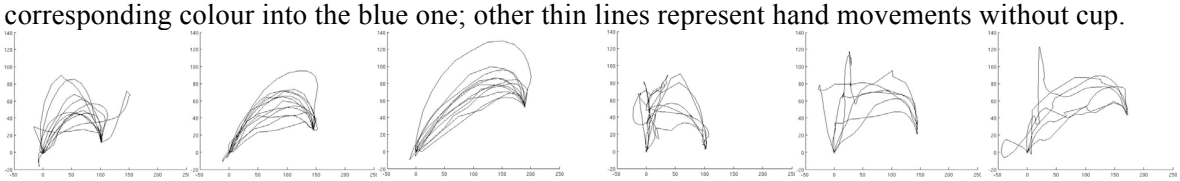


Fig. 2: Normalised hand trajectories of groups of participants. (a) Adult-Adult-Interaction. 1st, 2nd, 3rd action. **(b) Adult-Child-Interaction.** 1st, 2nd, 3rd action.

Considering these instances across our corpus, the presenters' hand trajectories in the AA-condition appear to have a relatively homegenous parabolic shape (Fig.2a). The trajectories in the AC-condition show more variation (Fig.2b): higher arches and modulations especially at the trajectory onset and the third action seems to be less pronounced. From an interactional point of view the question arises: Can we find any interactional reasons for the variability in the participants' hand trajectories? How are such hand trajectories instantiated in and through the unfolding course of the interaction? What function might they have for the participants and the action being carried out?

4 Trajectory Peaks: Organising the recipient's attention

Starting from the assumption that an adult's presentation of some task to his/her young child is a joint activity, invites us to understand the adult's presentation in relation to the recipient's actions. In our case, the action of stacking cups requires the recipient to attentively observe. Therefore, we focus in close detail on the interplay of the adult's presentation and the recipient's attention.

For detailed analysis, we consider the interaction which has led to the hand trajectories shown in Fig. 1e (VP001_1_AC_Becher): a 1st high arch (green), a 2nd curve with multiple peaks (yellow) and a 3rd rather flat trajectory (red). In this case, a father (VP) and his child (RC) are sitting face-to-face across a table, when the experimenter places the new set of toys on the table. VP looks at the child, brings his left hand towards the green cup, and the child immediately begins to shift his gaze to it. Then, VP begins to lift the cup, and once it arrives in mid-air (img.1), the child begins to raise his gaze following VP's hand/green cup until it comes to a halt (img.2). Once the child's gaze has reached his hand, the father begins to talk "LOOK; first of all we take the GREEN (one);" At the end of "GREEN (one)", VP begins to move his hand/cup down towards the blue one, and again the child's gaze follows, a few fragments of a second later, down to the blue cup (img.3). Thus, the sequential ordering of these actions – and the delay of the child's gaze with regard to VP's hand motion – suggests that the child uses the adult's hand motion as an orienting device for where to look at. At the same time, VP orients to the child's following, in that he delays the continuation of his own action (here: to start talking) until the appropriate focus of attention has been established.

About a second later, the child's gaze shifts from the blue/green cup to the side, while VP begins to grab the yellow cup with the other hand (img.4) – thus, child is looking in the wrong direction. When VP begins to lift the yellow cup – while also verbally announcing the new action – the child is still gazing to the other side. VP reacts by firstly stopping his hand motion, then shakes the cup while uttering "HELLO RASMUS; look HERE". This, in turn, engenders a shift in the child's orientation who begins to turn towards the yellow cup (img.5). Again, the adult's hand motion serves to re-orient the child's attention. Again, once the child's gaze visibly reaches VP's hand/ yellow cup, VP restarts his action. He lifts his hand/yellow cup further (img.6; this creates the second peak), however, at the same moment, the child's gaze fades down towards the blue cup. Two seconds later, VP finally drops the yellow cup into the blue one. As it turns out, the child appears to have *anticipated* the next step in the task. More complex than the first stacking action, this second one is visibly constituted of two seperate movements, the first of which serves as an *explicit* orienting device for the child.

Then, to initiate the third action, VP moves his right hand over to the red cup and, while he grabs it, the child's gaze immediately follows (img.7). When the child's line of sight reaches the red cup, VP moves it closer to the child (img.8) and also verbally starts the next action ("A:ND THEN,"). He then moves the red cup – in a distinctly *flat* trajectory – to the blue one. Again, the child's gaze follows immediately (img.9). Thus, it appears that under certain conditions – the child being oriented to the action and having displayed correct expectations about relevant next steps – parents re-adjust their previously *pronounced* hand motions (resulting in trajectories with high arches or peaks) to perform *flat* trajectories – similar to the ones know from adult-adult-interactions.

This case analysis reveals a direct relationship between the ways in which parents modify – step by step in the unfolding course of the interaction – their actions with regard to the child's focus of attention. Action modification and the recipient's gaze can be seen to have a reciprocal sequential relationship and constitute a constant loop of mutual adjustments: The presenter's hand motion helps to orient the child's attention while the child's line of sight is consequential for the concrete ways in which the action is performed. This way, we can identify some preliminary inter-dependencies between hand motions and interactional organisation (which we also find across the corpus): (a) high arches – engaging recipient to follow; (b) peaks – organizing attention; (c) flat trajectories – reacting on the recipient's display of correct expectations about relevant next actions.

5 Flat trajectories: Loosing the recipient's attention

Given our observations on the interplay of action modification and the organisation of interaction, a test scenario for their validity are further cases in which parents perform similar motions under

different conditions. For example: If pronounced hand motions indeed serve as orienting devices for the child and flat hand trajectories react upon the display of understanding (and thus are likely to occur rather in the 3rd stacking action), we could hypothesize that flat hand motions during the presentation would rather *not* engage the infant. In the video data, this could be seen either as the child not following the action demonstration by gazing appropriately or by disattending.

To test this, consider the interaction which has led to the series of flat hand motions shown in Fig. 1c (VP052_3_AC_Becher): When the toys are placed on the table, the child is gazing towards the floor. VP then verbally calls for the child's attention ("HANNA; HAVE a look"), the child reacts by looking to the cups, but then re-orients to the floor – just when VP grabs the green cup (01).

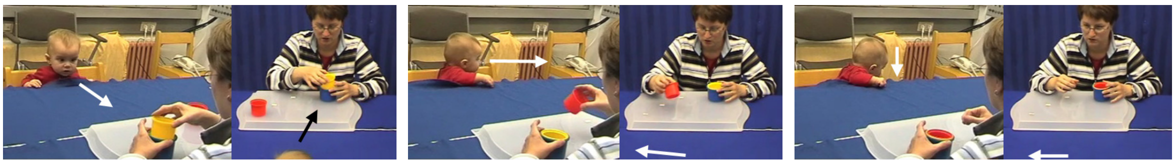
```
01 VP:      ↑HANna; KUCK mal; (0.2) ↑HAN|na; (0.5) |DA; die Becher kennst du auch; |ne,
   RC-gaz: @Ø                           |>>>>>>>>>>|@b                             |>>>
```

Again, VP verbally calls the child ("HANnah"), the child then re-orients to VP, VP notices that the child is paying attention and quickly grabs the green cup. She lifts it slightly (img.1) and moves it over to the blue one. With a little delay, the child's gaze follows to the blue cup (img.2).

```
02 VP:     |(0.5)|↑HANnah; (0.5) |kuck mal |↑HI:E|R; (0.2) den kann man da REIN |stel|len;
   VP-act: |to g |                         |grab g |g>bl                       |
   RC-gaz: |@Ø                  |>>>|@g (following)                                  |>>>>
                                            *1                                 *2
```

Then, while VP's right hand moves to the green cup, the child's gaze remains in the opposite direction. VP briefly gazes at the child, sees her slightly disattending, and – nevertheless – reorients to the cup and moves it straight to the blue one (img.3). Just at the moment when the yellow cup is being dropped into the blue one (img.4), the child reorients to it. Thus, the child has not attended to the actual action, but to VP, who only looks at her recipient at the end of the action, it appears as if the child is gazing correctly and attending to the action. VP, then procedes to grab the red cup and move it into the blue one with another *flat* motion. The child disattends immediately (img.5 and 6).

```
03 VP:     (0.2)|(0.2) und |DIEsen kann man da |rein |stellen; |(0.4) und den RO|ten;
   VP-act:      |y>b                           |               |r>b             |
   RC-gaz: @Ø              |@bl/y                    |>>>>>>>>>>|@Ø
                        *3       *4                              *5                *6
```



This conduct confirms our hypothesis and suggests that presentational actions without modification do not seem to be appropriate for helping the child to attent to the relevant events.

6 From single cases to the corpus: Linking qualitative and computational analysis

Basing on the qualitative analysis presented above, a next step consists in testing our findings across the entire corpus. While further case analyses support our findings, we have begun to develop ways of linking qualitative and quantiative methods: From the qualitative analysis we are able to derive a set of relevant aspects of interactional conduct, which we use as categories for systematic annotation: hand trajectories, gaze, talk. Using computational methods we can now investigate correlations between these interactional variables. The following visualisations (for the two cases analysed above) enable us to directly detect e.g. those moments, in which the child does (marked in green) and does not gaze to the appropriate location during the action presentation (marked in red), whether the tutor/adult is aware of the child's attending/disattending or not (dark vs. light), the child's anticipating gaze (blue) and how those instances link to the hand motion.

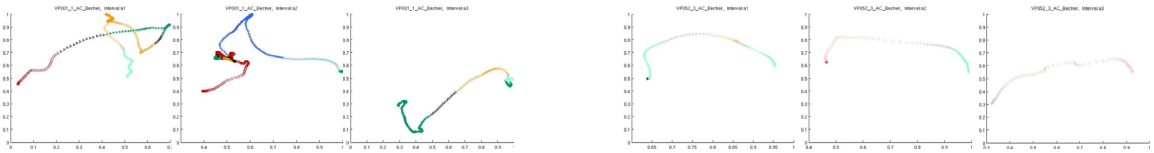


Fig.3: Motion trajectories and gaze. (a) VP001_1_AC (Fig. 1e). **(b)** VP052_3_AC (Fig. 1c)

7 Discussion and Implications

Starting from the observation made in developmental research that tutors modify both their speech and actions for the learner, we aimed at investigating – from an interactional point of view – the sources of its variability. We have been able to show a direct relationship between the ways in which parents modify – step by step in the unfolding course of the interaction – their actions with regard to the child's focus of attention. Action modification and the recipient's gaze can be seen to have a reciprocal sequential relationship and constitute a constant loop of mutual adjustments: The presenter's hand motion helps to orient the child's attention while the child's line of sight is consequential for the concrete ways in which the action is performed. This way, we were able to identify some preliminary interdependencies between hand motions and interactional organisation: (a) high arches – engaging recipient to follow; (b) peaks – organizing attention; (c) flat trajectories – reacting on the recipient's display of correct expectations about relevant next actions.

These results have implications with regard to: (a) *Social Learning*: Our results support the view that, within a social exchange, information emerges from and within the communicational process (Fogel & Gravey 2007) as opposed to approaches arguing for individualistic representation formation. (b) *Conceptualization of Gestures*: Albeit our analysis focuses on one particular example of manual actions, it yields to a more general topic in gesture research: "Gestures" have generally been considered as one single unit that participants deploy in interaction, eventually copy and modify from their co-participants. However, we argue for a dynamic concept of gestures considering the coparticipant's conduct as consequential for the concrete appearance, shape and timing of gestures. (c) *Methodological issues*: As gestures are per se fugitive motions, a general concern for gesture research consists in how to best track and make visible hand movements and trajectories. We suggest that 2D hand tracking using computational pattern recognition methods can be a valuable tool that can be used off-line on existing video data. This way, research would not necessarily be restricted to laboratory settings with sophisticated motion capture systems.

Our future work will extend the analysis towards quantitative issues using computational methods.

Acknowledgment

Karola Pitsch, Katharina Rohlfing, and Britta Wrede gratefully acknowledge the financial support from the FP7 European Project ITALK (ICT-214668) and Anna-Lisa Vollmer the financial support from Honda Research Institute Europe for the project 'Acquiring and Utilizing Correlation Patterns across Multiple Input Modalities for Developmental Learning'.

Bibliography

Brand, Rebecca J, Dare A Baldwin, and Leslie A Ashburn. "Evidence for 'Motionese': Modifications in Mother's Infant-Directed Actions." Developmental science 5, no. 1 (2002): 72-83.

Rohlfing, Katharina, Jannik Fritsch, Britta Wrede, and Tanja Jungmann. "How Can Multimodal Cues From Child-Directed Interaction Reduce Learning Complexity in Robots?" Advanced Robotics 20, no. 10 (2006): 1183-199.

Fogel, A. and Garvey, A. (2007): Alive communication. In: Infant Behavior and Development 30: 251-257.

Gogate, L J, L E Bahrick, and J D Watson. "A Study of Multimodal Motherese: The Role of Temporal Synchrony Between Verbal Labels and Gestures." Child development (2000): 878-894.

Mondada, Lorenza. "Participants' Online Analysis and Multimodal Practices: Projecting the End of the Turn and Closing of the Sequence." Discourse Studies (Special Issue: Discourse, Interaction and Cognition) 8, no. 1 (2006): 117-129.

Pitsch, Karola. Sprache, Körper, Intermediäre Objekte: Zur Multimodalität der Interaktion im Bilingualen Geschichtsunterricht. PhD dissertation, Bielefeld University (to appear in the series "Qualitative Soziologie"). 2006.

Vollmer, Anna-Lisa, Katrin Solveig Lohan, Kerstin Fischer, Yukie Nagai, Karola Pitsch, Yannik Fritsch, Katharina J. Rohlfing, and Britta Wrede. "People Modify Their Tutoring Behavior in Robot-Directed Interaction for Action Learning." In ICDL 2009. 2009.

Zukow-Goldring, P. "Sensitive Caregiving Fosters the Comprehension of Speech: When Gestures Speak Louder Than Words." Early Development and Parenting 5, no. 4 (1996).