Grant Agreement no. 214668

ITALK
Integration and Transfer of Action and Language Knowledge in robots

Seventh Framework Programme (FP7)
Cognitive Systems, Interaction, Robotics (ICT-2007.2.1)

Internal Report 3.2
Report on analysis of parent-child interaction on object knowledge demonstration and specifications of feature set for detecting tutoring behaviour as a basis for a “tutor spotter”

Deliverable Submission: Month 24

Project Start Date: 2008-03-01  Duration: 48 months
INTERNAL REPORT 3.2 – TABLE OF CONTENT

1. Introduction 3
2. Analysis of parent-child interactions 3
3. Feature set for a tutoring spotter 5
4. List of papers 5
5. References 6
6. Papers
Report on analysis of parent-child interaction on object knowledge demonstration and specifications of feature set for detecting tutoring behavior as a basis for a “tutor spotter”

1. Introduction

This report focuses on the analysis of features that may help to detect tutoring behavior, as proposed by Csibra & Gergeley (2005) in parent-child interactions. The data underlying this project is based on 65 parents-child dyads, in which the task for the parents was to demonstrate 10 objects and their functions to their children. The analyses were carried out based on quantitative and qualitative methods using computational tracking systems of human motion developed by Fritsch et al. (2005) as well as speech based analysis tools. The goal is to derive feature sets that are suitable to detect tutoring behavior in an interaction partner that can be used to implement a “tutoring spotter”. Such a spotter will on the one hand make a system sensitive to learning situation. On the other hand, it is argued that in a learning situation, some scaffolding methods are applied by the tutor according to which more structured information is provided to the learner. The project thus continues on multimodal analysis of the collected data at different levels of granularity in order to find out on which levels the variability of the tutoring behavior occurs. Based on these results further experiments from parent-infant or tutor-robot interactions are needed. The tutoring spotter will be evaluated by integrating it into a robot system (implemented on iCub) where the effect of the robot’s reaction (i.e. signalling attention) upon the detection of tutoring behavior on the behavior of the tutor will be analyzed in more details.

2. Analysis of parent child interaction

The starting point of our analyses is the cross-sectional ‘motionese study’, in which adult-child interactions were investigated in comparison to adult-adult interactions. Our dependent measures were motionese and eye gazing behavior. The term “motionese” accounts for modified action demonstration towards children (Brand et al., 2002; Rohlfing et al., 2006). In the field of developmental robotics, research often assumes that in human robot interaction (HRI), robots are treated similar to infants, because of their immature apparent cognitive capabilities. Concerning this hypothesis we conducted a study – similar to the motionese study – with a simulated robot called Ackachan, which was equipped with a visual attention system based on the idea of Itti and Koch (1998) (see Figure: 2). In Vollmer and Lohan et al. (2009A), we were able to show that there are significant differences in hand movement velocity, motion pauses, range of motion, and eye-gaze suggesting that for example adults decrease their hand movement velocity in an Adult-Child Interaction (ACI), opposed to an Adult-Adult Interaction (AAI) and this decrease is even higher in the

Figure 1: 4 different age groups are participated in the parent-infant study.
Adult-Robot Interaction (ARI). We also found important differences between ACI and ARI with respect to how the behavior is modified over time as the interaction unfolds.

Regarding these findings, we studied the learners’ differences with respect to their different age to get a closer look on behavior modification of a tutor influenced by child’s development. Our data suggest that actions chosen to attract attention can primarily be found in interaction with younger infants, whose attention needs more guidance. Interactions with older children seem to differ due to either the increase of children's attention abilities or that parents use other means to attract their attention. In contrast, parameters that appear to be more in charge of structuring the action seem to persist over the children's age and their verbal capabilities, shown in “Which ‘Motionese’ Parameters Change With Children's Age?” by Vollmer and Lohan et al. (2009, B) (see Figure: 1).

However, we found that also the eye gazing behavior is a very precise indication for the acceptance of an interaction partner and could give us a hint on the quality of a tutoring situation (Lohan and Vollmer et al., 2009). Infants seem to be sensitive to tutoring situations and they detect these by ostensive cues (Csibra & Gergely, 2005). Csibra and Gergely argued that contingency is a characteristic ostensive stimulus of a tutoring situation. Contingency could be measured in respect to the eye gazing behavior of a tutor (Lohan & Vollmer et al., 2009). The ostensive signals considered here appear practical for the robot to detect situations in which it is being tutored, but we argue that a robot cannot make use of an important ostensive stimulus such as contingency without providing the “right” signals for the interactional construct. More specifically, we found that the eye-gaze behavior in the ARI situation is rather similar to that of the AAI situation, with less time of the eye-gaze being spent on the interaction partner. This is congruent with previous findings from Vollmer and Lohan et al. 2009A. Assuming that eye-gaze is also used in order to check for understanding of the partner, the eye-gaze behavior directly after the end of a task becomes relevant. Indeed, we can see that the eye-gaze lengths in pauses after a performed sub task is significantly longer in ACI as opposed to AAI. Thus, the parents appear to look for understanding in their infants. Interestingly, the behavior in ARI tends to be similar to the one in AAI indicating that adults behave differently towards robots.

Illustration 1: Figure 2: The Ackachan study. a) learner b) tutor c) setup

Handtrajectories

Eye-Gaze to the interaction partner

Figure 3: Illustrated main results of our Ackachan Experiment, all results are based on the analysis of the tutor's behavior.
These findings indicate the necessity of integrating top-down feedback structures into a bottom-up system for robots to be fully accepted as interaction partners.

One possible explanation for the impaired eye gazing behavior in the simulated robot situation (cf. Figure 3) might be the fact that it was difficult for the tutor to understand where the robot was looking at because of missing 3D information from the face and the eyes. We therefore conducted a third study with the physically embodied robot iCub. The iCub was equipped with the same visual attention system as the Ackachan simulation before and the experiment was carried out in analogy to the experiment with Ackachan. Moreover, we were able to recruit the same participants, which fact allowed us to carry out within-subject analyses. The iCub was equipped with two different eye-gaze movements: in condition (1) it only moved the eyes to the most salient point in the image, this is similar to Ackachan's behavior; in condition (2) the eye movement was followed by a head turn. This way we hypothesized the eye-gaze became more salient to the tutor and would induce a more social experience. For the results see Lohan, et al. (2010 under preparation).

3. Feature set for a tutoring spotter

With respect to detecting a tutoring situation, our current findings suggest that the tutor needs to be motivated in order to accept a robot as a social interaction partner and to get a polemical attitude of a tutor. This polemical attitude is much easier to detect as it can be measured using motionese and contingency features. It is necessary to monitor the behavior of the tutor for the purpose of determining when a tutoring situation begins in time. Cross-modal cues such as speech, beep, motion, motionese, gaze, still phase could contribute to the monitor process. However, the perception of these cues is not sufficient for a tutoring situation. In addition, it is necessary to perceive them within the variability of human behavior. We believe that the variability in behavior could be manipulated by the given feedback of our system and this could affect the way, in which the information will be packaged. However, we still do not know which feedback cues lead to which change in the variability of the tutor behavior. Previous finding reveal that the actions of the learner and the tutor's presentation are closely intertwined, as has been shown for children's eye gazing behavior and the concrete ways in which parent's hand movements are carried out during action demonstration (Pitsch, et al., 2009). Assuming that there is an important role of the timing, this lead us to the concept of temporal contingency as a feedback strategy and a measure of the quality of the interaction.

4. Papers

The following papers are attached:

1. Lohan, K. et al. (2009)
References


Analyzing Contingency in Tutoring Situations
Katrin Solveig Lohan, Katharina Rohlfing, and Britta Wrede

Colt-Lab, Bielefeld University, Germany. Applied Informatics Group, Bielefeld University, Germany.
EU-Project ITALK

HUROBINT@MIT 2009

Abstract

In developmental research, tutoring behavior has been identified as scaffolding infants’ learning processes. It has been defined in terms of child-directed speech (Motherese), child-directed motion (Motionese), and Contingency. Contingency describes situations in which two agents socially interact with each other and Cabra and Gergely showed that contingency is a characteristic aspect of social interaction [3].

In the field of developmental robotics, research often assumes that in human-robot interaction (HRI), robots are treated similar to infants, because their immature cognitive capabilities benefit from this behavior. Here we present results concerning the acceptance of a robotic agent in a social learning scenario obtained via comparison to adults and 8-11 months old infants in equal conditions. These results constitute an important empirical basis for making use of tutoring behavior in social robotics.

Results

Concerning action modification we replicated the results achieved in [2], [1] and [6]. Concerning contingency we found:

- ACI eye-gaze bouts were most frequent to interaction partner
- frequency of eye-gaze bouts to object is significantly lower in ARI
- average significantly longer bouts in ACI toward an interaction partner
- average length of eye-gaze bout to object was significantly smaller for ACI
- In ACI significantly more time was spent gazing at the interaction partner
- total length of eye-gaze bouts to object is significantly lower in ACI

⇒ No contingent eye-gaze patterns in ARI
⇒ Outlook: How to elicit contingency in ARI?

Two Experiments

- adult-child interaction (ACI) and adult-adult interaction (AAI) like in [6] and [5]
- adult-robot interaction (ARI) using the example of a robot simulation equipped with a bottom-up saliency-based attention model [5]

Research Questions

J.S. Watson thinks of contingency as the human infant’s means for detecting socially responsive agents and therefore postulates the existence of an innate contingency detection module as one of the most fundamental innate modules [7].

"The discovery that another agent’s gaze is a cue worthy of monitoring relies on the infant’s ability to detect the contingency structure in interactions with that agent" [4].

Contingency detection (‘production’) as an important prerequisite for a ‘developing’ robot

References


Acknowledgment & References

Katrin Solveig Lohan, Katharina Rohlfing, and Britta Wrede gratefully acknowledge the financial support from the FP7 European Project ITALK (ICT-216446).

Contact: klohan@techfak.uni-bielefeld.de

http://www.cor-lab.de/corlab/cms/node/139
http://aiweb.techfak.uni-bielefeld.de
Which ostensive stimuli can be used for a robot to detect and maintain tutoring situations?

Katrin Solveig Lohan  Anna-Lisa Vollmer  Jannik Fritsch  Katharina Rohlfing  Britta Wrede  
CoR-Lab, Applied Informatics Group  Bielefeld University, Bielefeld, Germany  
http://www.cor-lab.de  
klohan@techfak.uni-bielefeld.de

Abstract

In developmental research, tutoring behavior has been identified as scaffolding infants’ learning processes. Infants seem sensitive to tutoring situations and they detect these by ostensive cues [4]. Some social signals such as eye-gaze, child-directed speech (Motherese), child-directed motion (Motionese), and contingency have been shown to serve as ostensive cues. The concept of contingency describes exchanges in which two agents interact with each other reciprocally. Csibra and Gergely argued that contingency is a characteristic ostensive stimulus of a tutoring situation [4]. In order for a robot to be treated similar to an infant, it has to both, be sensitive to the ostensive stimuli on the one hand and induce tutoring behavior by its feedback about its capabilities on the other hand.

In this paper, we raise the question whether a robot can be treated similar to an infant in an interaction. We present results concerning the acceptance of a robotic agent in a social learning scenario, which we obtained via comparison to interactions with 8-11 months old infants and adults in equal conditions. We applied measurements for motion modifications (Motionese) and eye-gaze behavior. Our results reveal significant differences between Adult-Child Interaction (ACI), Adult-Adult Interaction (AAI) and Adult-Robot Interaction (ARI) suggesting that in ARI, robot-directed tutoring behavior is even more accentuated in terms of Motionese, but contingent responsivity is impaired. Our results confirm previous findings [14] concerning the differences between ACI, AAI, and ARI and constitute an important empirical basis for making use of ostensive stimuli as social signals for tutoring behavior in social robotics.

1. Introduction

In social learning, infants benefit from the behavior of their tutors. The modified behavior seems to help infants to filter the information that is crucial for learning. Csibra and Gergely [4] highlight the importance of this pedagogic behavior that is crucial for the understanding of some actions: "pedagogy essentially created a new way of information transfer among individuals through the use of ostensive communication". In their work, they give the example of peeling a hard fruit or carve away pieces of wood with a tool. The movement and the tool in both actions are the same, but the goal and reason for the action are very different. Where it is easy to infer the goal of the action when peeling a fruit, i.e., getting to the edible parts, it is not obvious what is intended in the case of the wood carving. Therefore, tutoring is crucial in order for a learner to understand the goal correctly. Csibra and Gergely [4] argue that economical reasons account for tutoring, because otherwise learning would not be feasible. Tutoring situations thus are created by the tutor via ostensive stimuli, which are "originally evolved to assist pedagogy". The effect of pedagogy seems to rely on the bidirectionality. Csibra and Gergely (2005) explain the contribution achieved by the learner, who has to send signals during the course of tutoring telling the tutor when s/he is attentive and receptive and possibly showing understanding. Furthermore, infants seem sensitive to tutoring situations and ostensive cues help them to detect these [13]. The term "ostensive cues" refers to social signals such as eye-gaze, child-directed speech (Motherese) [5], child-directed motion (Motionese) [2,6,7], and contingency [4]. While the phenomenon of multimodal child-directed speech (Motherese) or action (Motionese) is widely known, the concept of contingency is less popular. It describes exchanges in which two agents interact with each other reciprocally. Csibra and Gergely ([4], p.8) argue that contingent responsivity is a characteristic ostensive stimulus of a tutoring situation: "If a source repeatedly appears to remain silent during your actions but starts to emit signals as soon as you have stopped your actions, it gives
you the strong impression that the source is communicating with you”. The idea of creating a robot that actively filters the information from the environment and manages to attend to certain sources of information while ignoring others has to be supported by the robot’s sensitivity to the ostensive stimuli on the one hand and induce tutoring behavior by its feedback about its capabilities on the other hand. A robot which has the appearance of an infant should hence be able to profit from these behavior modifications as well. Recently, Vollmer et al. found that adults modify their behavior when interacting with children (ACI) and robots (ARI) as opposed to adult-directed interaction (AAI) [14]. Modifications were found with respect to Motionese measurements, indicating that in ACI and ARI movements were slower, less round and had a slower pace than in AAI indicating that subjects behave similar towards robots and infants. However, number and length of eye-gaze bouts differed significantly between ACI and ARI with less eye-gaze bouts and less long eye-gaze bouts directed towards the interaction partner in ARI. This indicates that contingency was impaired in the ARI condition. In this paper, we report on results from a task with a similar structure based on a more fine grained analysis of the eye-gaze behavior in order to

- show how far the findings by Vollmer et al. hold for a different task
- analyze the structure of eye-gaze behavior over time
- discuss these results with respect to the question in how far the observed modifications of behavior can be interpreted as ostensive signals in human-robot interaction.

2. Experiment

Two experiments were carried out to obtain data from parent-infant and adult-robot interactions [14]. The data on adult-child interaction is based on the same setting as in [12] and [10]. The data on human-robot interaction was obtained in a second experiment as described in [14]. From the overall set of items that were presented we selected the "Minihausen” task. This task is similar to the stacking-cups task as it is a rather goal-directed action with three sub-goals to be reached. Results from analyses of motionese and contingency features in parent-infant and adult-robot interaction have shown that while motionese features of infant-directed and robot-directed interactions are similar, they diverge for contingency measures, indicating that contingency is impaired in human-robot-interaction, [14]. In this paper we ask the question in how far these results are decisive for the statement that motionese as well as contingency features serve the function of ostensive signals.

2.1. Motionese Experiment (ME)

2.1.1 Subjects

The Motionese Corpus consists of infant- and adult-directed interactions. We selected the younger group comprising 12 families of 8 to 11 months old children. Both parents were asked to demonstrate functions of 10 different objects to their children as well as to their partners or another adult. In the following, we focus on the analysis of the "Minihausen” task, because it offers good comparability in motion performance. We further selected a subgroup of 8 parents (4 fathers and 4 mothers) for the ACI and a subgroup of 12 parents (7 fathers and 5 mothers) for the AAI, because of the quality of the video, sound and due to the way in which the action was performed. More specifically, the order in which the blocks of the considered "Minihausen” task are put onto the wooden base poles can vary: We selected only those parents, who started the task by putting the first block -the one closest to the body- onto the respective pole which means putting the blue block onto the rightmost pole. (see Fig. 3 a1).

2.1.2 Setting

Parents were instructed to demonstrate a ”Minihausen” task to an interaction partner. The interaction partner was first their infant and then an adult. Fig. 1 illustrates the top-view of the experimental setup. The ”Minihausen” task was to sequentially pick up the blue (a1), the yellow (a2), and the green (a3) block and put them onto the wooden base with three poles on the white tray.

![Figure 1: Motionese Setting](image)

2.2. Robot-Directed Interaction Experiment (RDIE)

2.2.1 Subjects

31 adults (14 females and 17 male) participated in this experiment 7 out of which were parents as well. Out of this group, we selected 12 participants (8 female and 4 male), who performed the task in a comparable manner.
### 2.2.2 Setting

The participants were instructed to demonstrate several objects to an interaction partner, while explaining him/her how to do it (Fig. 2). Again we chose the "Minihausen" task for analysis. The interaction partner was an infant-like looking virtual robot with a saliency-based visual attention system [10]. The robot-eyes will follow the most salient point in the scene, which is computed by color, movement, and other features (see [10]).

![Figure 2. The robot simulation presented on the screen can be seen on the left picture. The right picture shows the Robot-directed Interaction Setting, there are four cameras which are recording the scene. The subject is seated across from the robot and the object is laid on the table in front of the tutor.](image)

### 3. Data Analysis

The goal of this paper was to analyze those cues, that we hypothesize to serve as social signals in tutoring behavior. These can be grouped into two groups, one that measures Motionese and another one that may be used to measure Contingency. We coded the videos semi-automatically to obtain data for the 2D hand trajectories and the eye gaze directions.

#### 3.1. Annotations

For all annotations, we used the video captured by camera (cam) 1, see Fig. 1 and 2. It shows the front view on the demonstrator and is therefore best suited for action, movement, and gaze annotations, which are discussed in detail below.

![Figure 3. The action was devided into movement and pause parts and into subactions. This graphic shows an example for the structure of an 'Action', 'Subaction'(intro = Introduction and sum = summary), and 'Movement'.](image)

### 3.2. Measures

For quantifying Motionese and Contingency, we computed five variables related to the 2D hand trajectories derived from the videos and the eye gaze bout annotations produced with Interact.

#### 3.2.1 Motionese

**Action Segmentation:** For analyzing the data, the action of the "Minihausen" task and additionally, the sub-actions (a1-a3) of grasping one block until releasing it onto the end position (Fig. 3) were marked in the video. We defined

1. action as the whole process of transporting all objects to their goal positions.
2. subaction as the process of transporting one object to its goal position.
3. movement as phases where the velocity of the hand is above a certain threshold. All other phases are defined as pauses.

**Hand Trajectories:** The videos of the two experiments were analyzed via a semiautomatic hand tracker system (Fig. 4). The system is written as a plug-in for a graphical plug in shell, iceWing [8], and makes it possible to track both hands with an Optical Flow based algorithm, Lucas & Kanade [9]. It allows manual adjustment in case of tracking deviation. We used this tracking system instead of a previously used 3D body model system, [12], since 3D results in [12] were not significant, we focused on 2D analyses which provide to show more stable results. Additionally, the new system is easily accessible for non-expert users.

![Figure 4. Example frame for hand tracker system annotation. The red and violet circles depict the tracking regions. The points in the middle of the circles are the resulting 2D points for the hand trajectory.](image)

#### 3.2.2 Contingency

**Eye Gaze:** In annotating the eye gaze directions with the program Interact [1], we distinguished between looking at the interaction partner, looking at the object and looking anywhere else.

### 3.1. Motionese

**Action Segmentation:** For analyzing the data, the action of the "Minihausen" task and additionally, the sub-actions (a1-a3) of grasping one block until releasing it onto the end position (Fig. 3) were marked in the video. We defined

1. action as the whole process of transporting all objects to their goal positions.
2. subaction as the process of transporting one object to its goal position.
3. movement as phases where the velocity of the hand is above a certain threshold. All other phases are defined as pauses.

**Hand Trajectories:** The videos of the two experiments were analyzed via a semiautomatic hand tracker system (Fig. 4). The system is written as a plug-in for a graphical plug in shell, iceWing [8], and makes it possible to track both hands with an Optical Flow based algorithm, Lucas & Kanade [9]. It allows manual adjustment in case of tracking deviation. We used this tracking system instead of a previously used 3D body model system, [12], since 3D results in [12] were not significant, we focused on 2D analyses which provide to show more stable results. Additionally, the new system is easily accessible for non-expert users.

![Figure 4. Example frame for hand tracker system annotation. The red and violet circles depict the tracking regions. The points in the middle of the circles are the resulting 2D points for the hand trajectory.](image)
the action per frame as the average velocity for subactions a1, a2, and a3 each.

*Range* was defined for each subaction separately as the covered motion path divided by the distance between motion, i.e. subaction, on- and offset.

### 3.2.2 Contingency

The Contingency of the interactions was quantified in terms of variables related to eye gaze, as defined in [3] for measuring interactivity.

The *total length of eye-gaze bouts to interaction partner* defined as the percentage of time of the action spent gazing at the interaction partner was computed. Brand et al. found that the total length of eye-gaze bouts to the interaction partner in their study was significantly greater in ACI than in AAI [3]. Also the *total length of eye-gaze bouts to object* and the *total length of eye-gaze bouts elsewhere* were calculated as the percentage of time of the action spent gazing at the object and somewhere else, as for example at the table or the experimenter.

### 4. Results

A non-parametric test (Mann-Whitney U test) was run for all pairs of samples, ACI vs. AAI, ACI vs. ARI, and AAI vs. ARI. Table 1 depicts the results of the study.

#### 4.1. Motionese

For the Motionese measures, our results revealed the following:

For the *velocity* measure, which is computed for each subaction and takes into account the hand movement during the transportation of the respective block, the results showed significant differences for all three subactions for all pairs of conditions. These results clearly show that in AAI hand movements are faster than in ACI and ARI and additionally that hand movement is slowest in the ARI condition. Also note that for all conditions the mean values increase for the consecutive subactions: velocity in subaction a1 < velocity in a2 < velocity in a3. In ARI, the rate in which the mean values increase is lowest and in AAI the rate is highest. The latter is specially noticeable for the last subaction a3. The *range* measure suggests that ARI exhibits the greatest range for each subaction and therefore movement is most exaggerated. Also, range is greater in ACI than in AAI. For ACI vs. AAI results revealed no significance, but a trend for subactions a2 and a3. For ACI vs. ARI solely results for subaction a3 showed significance, for a1 and a2 they show a trend. For AAI vs. ARI subactions a2 and a3 revealed significance, whereas a1 again shows a trend. Again we can state that in ARI the first subaction a1 has the highest range value of all subactions over all conditions. Looking at this measure over time, range decreases rapidly to about one half for subaction a2 and some more for the last subaction a3. For the other conditions however the rate of change, i.e. the decrease, is not as drastic.

![Figure 5](image-url) This graph shows the range of hand movement in the three different subactions on the left. On the right, the mean velocity of hand movement in the three different subactions can be seen for the "Minihausen"-task (y-axis) in every condition (x-axis).

#### 4.2. Contingency

Most interestingly, the results for eye gaze show a completely different picture. For *total length of eye-gaze bouts to interaction partner* they show that in ACI significantly more time was spent gazing at the interaction partner than in AAI and ARI. Differences between AAI and ARI are not significant. Looking at this measure over time, it is interesting to notice that in all three conditions the most time of gazing at the interaction partner was spent in the summary part of the action, sum.

For the measure *total length of eye-gaze bouts to object*, values are significantly lower in ACI than in AAI and ARI, where differences between AAI and ARI exhibit that values are significantly lower in ARI.

The *total length of eye-gaze bouts elsewhere*, which measures the percentage of time gazed neither to interaction partner nor object, reveals that most time gazing somewhere else is spent in the ARI condition, followed by ACI. The differences between ACI and AAI could be a result of the design of the study, because the AAI follows the ACI, so that instructions and experimenter are not anymore needed to turn to for help in the demonstration of the task, because it has already been shown once. Additionally, in all conditions it is gazed elsewhere mostly in p1 and p2 and not during the transportation of the cups in a1, a2 and a3.

### 5. Conclusion

To conclude, we did find ostensive signals in tutoring situations in adult-robot interaction. On the one hand, our results for range and velocity show significantly exaggerated
<table>
<thead>
<tr>
<th>Variable</th>
<th>ACI M</th>
<th>ACI SD</th>
<th>AAI M</th>
<th>AAI SD</th>
<th>ARI M</th>
<th>ARI SD</th>
<th>Z</th>
<th>Z</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>velocity a1</td>
<td>3.58</td>
<td>0.81</td>
<td>4.72</td>
<td>1.39</td>
<td>2.08</td>
<td>0.86</td>
<td>-2.394**</td>
<td>-3.666***</td>
<td>-3.747***</td>
</tr>
<tr>
<td>velocity a2</td>
<td>4.19</td>
<td>1.84</td>
<td>6.39</td>
<td>1.71</td>
<td>2.59</td>
<td>0.87</td>
<td>-2.535**</td>
<td>-2.792**</td>
<td>-3.982***</td>
</tr>
<tr>
<td>velocity a3</td>
<td>6.62</td>
<td>2.43</td>
<td>11.78</td>
<td>2.95</td>
<td>3.73</td>
<td>1.51</td>
<td>-3.098***</td>
<td>-2.956**</td>
<td>-3.982***</td>
</tr>
<tr>
<td>range a1</td>
<td>4.22</td>
<td>2.49</td>
<td>3.41</td>
<td>0.72</td>
<td>6.29</td>
<td>5.53</td>
<td>-0.211</td>
<td>-1.369+</td>
<td>-1.288+</td>
</tr>
<tr>
<td>range a2</td>
<td>2.19</td>
<td>0.48</td>
<td>1.88</td>
<td>0.25</td>
<td>2.72</td>
<td>0.97</td>
<td>-1.549+</td>
<td>-1.314+</td>
<td>-2.635**</td>
</tr>
<tr>
<td>range a3</td>
<td>1.57</td>
<td>0.37</td>
<td>1.35</td>
<td>0.09</td>
<td>2</td>
<td>0.56</td>
<td>-1.479+</td>
<td>-2.409**</td>
<td>-3.396***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>ACI M</th>
<th>ACI SD</th>
<th>AAI M</th>
<th>AAI SD</th>
<th>ARI M</th>
<th>ARI SD</th>
<th>Z</th>
<th>Z</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>total length eye-gaze to i.p. in</td>
<td>10.86</td>
<td>14.52</td>
<td>6.65</td>
<td>7.15</td>
<td>6.65</td>
<td>7.15</td>
<td>-0.833</td>
<td>-1.419+</td>
<td>-0.76</td>
</tr>
<tr>
<td>total length eye-gaze to i.p. a1</td>
<td>27.81</td>
<td>25.02</td>
<td>9.01</td>
<td>16.92</td>
<td>9.25</td>
<td>11.38</td>
<td>-2.2*</td>
<td>-1.882*</td>
<td>-0.97</td>
</tr>
<tr>
<td>total length eye-gaze to i.p. p1</td>
<td>24.19</td>
<td>28.17</td>
<td>3.7</td>
<td>9.71</td>
<td>7.35</td>
<td>8.78</td>
<td>-1.853*</td>
<td>-1.03</td>
<td>-1.634+</td>
</tr>
<tr>
<td>total length eye-gaze to i.p. a2</td>
<td>15.39</td>
<td>16.67</td>
<td>2.42</td>
<td>4.44</td>
<td>3.16</td>
<td>4.81</td>
<td>-2.054*</td>
<td>-2.066*</td>
<td>-0.244</td>
</tr>
<tr>
<td>total length eye-gaze to i.p. p2</td>
<td>33.73</td>
<td>24.63</td>
<td>2.61</td>
<td>7.09</td>
<td>2.69</td>
<td>5.9</td>
<td>-3.055***</td>
<td>-3.306***</td>
<td>-0.082</td>
</tr>
<tr>
<td>total length eye-gaze to i.p. a3</td>
<td>23.05</td>
<td>23.09</td>
<td>4.37</td>
<td>8.71</td>
<td>6.2</td>
<td>10.48</td>
<td>-2.273*</td>
<td>-2.292*</td>
<td>-0.384</td>
</tr>
<tr>
<td>total length eye-gaze to i.p. su</td>
<td>43.8</td>
<td>23.81</td>
<td>27.55</td>
<td>7.43</td>
<td>19.66</td>
<td>13.65</td>
<td>-0.493</td>
<td>-2.793**</td>
<td>-1.878+</td>
</tr>
<tr>
<td>total length eye-gaze to o. in</td>
<td>69.29</td>
<td>29.43</td>
<td>82.32</td>
<td>22.47</td>
<td>62.65</td>
<td>8.7</td>
<td>-1.353+</td>
<td>-1.15</td>
<td>-2.817**</td>
</tr>
<tr>
<td>total length eye-gaze to o. a1</td>
<td>70.94</td>
<td>22.72</td>
<td>89.52</td>
<td>16.69</td>
<td>83.21</td>
<td>13.46</td>
<td>-2.1*</td>
<td>-1.213</td>
<td>-1.155</td>
</tr>
<tr>
<td>total length eye-gaze to o. p1</td>
<td>60.95</td>
<td>26.97</td>
<td>88.99</td>
<td>23.87</td>
<td>68.36</td>
<td>25.95</td>
<td>-2.273*</td>
<td>-0.714</td>
<td>-2.097*</td>
</tr>
<tr>
<td>total length eye-gaze to o. a2</td>
<td>82.68</td>
<td>18.18</td>
<td>96.2</td>
<td>8.19</td>
<td>92.43</td>
<td>7.85</td>
<td>-2.198*</td>
<td>-1.308+</td>
<td>-1.533+</td>
</tr>
<tr>
<td>total length eye-gaze to o. p2</td>
<td>65.02</td>
<td>25.55</td>
<td>97.39</td>
<td>7.09</td>
<td>80.23</td>
<td>22.36</td>
<td>-3.055***</td>
<td>-1.503+</td>
<td>-2.092*</td>
</tr>
<tr>
<td>total length eye-gaze to o. a3</td>
<td>76.95</td>
<td>23.25</td>
<td>95.63</td>
<td>8.71</td>
<td>87.23</td>
<td>13.77</td>
<td>-2.273*</td>
<td>-1.252</td>
<td>-1.721*</td>
</tr>
<tr>
<td>total length eye-gaze to o. su</td>
<td>55.79</td>
<td>22.63</td>
<td>52.71</td>
<td>31.88</td>
<td>57.92</td>
<td>17.94</td>
<td>-0.352</td>
<td>-0.109</td>
<td>-0.527</td>
</tr>
<tr>
<td>total length eye-gaze e. in</td>
<td>20.89</td>
<td>29.12</td>
<td>11.03</td>
<td>18.15</td>
<td>34.93</td>
<td>9</td>
<td>-0.624</td>
<td>-1.984*</td>
<td>-3.127***</td>
</tr>
<tr>
<td>total length eye-gaze e. a1</td>
<td>1.91</td>
<td>4.75</td>
<td>1.48</td>
<td>4.67</td>
<td>7.53</td>
<td>10.61</td>
<td>-0.52</td>
<td>-1.625+</td>
<td>-1.919*</td>
</tr>
<tr>
<td>total length eye-gaze e. p1</td>
<td>16.09</td>
<td>19.93</td>
<td>7.32</td>
<td>23.14</td>
<td>24.29</td>
<td>26.94</td>
<td>-1.501+</td>
<td>-0.812</td>
<td>-1.952*</td>
</tr>
<tr>
<td>total length eye-gaze e. a2</td>
<td>2.51</td>
<td>3.9</td>
<td>1.37</td>
<td>4.34</td>
<td>4.41</td>
<td>7.42</td>
<td>-1.178</td>
<td>-0.371</td>
<td>-1.604+</td>
</tr>
<tr>
<td>total length eye-gaze e. p2</td>
<td>2.38</td>
<td>5.35</td>
<td>0</td>
<td>0</td>
<td>17.08</td>
<td>20.59</td>
<td>-1.382+</td>
<td>-1.879*</td>
<td>-2.551**</td>
</tr>
<tr>
<td>total length eye-gaze e. a3</td>
<td>0.74</td>
<td>1.67</td>
<td>0</td>
<td>0</td>
<td>6.57</td>
<td>12.94</td>
<td>-1.382+</td>
<td>-0.877</td>
<td>-1.803*</td>
</tr>
<tr>
<td>total length eye-gaze e. su</td>
<td>1.09</td>
<td>2.31</td>
<td>7.65</td>
<td>11.74</td>
<td>22.42</td>
<td>15.92</td>
<td>-1.091</td>
<td>-3.507***</td>
<td>-2.267*</td>
</tr>
</tbody>
</table>

Table 1. Results of Mean, Standard deviation, Mann-Whitney U test, \( p < 0.1, \) \( * p < 0.05, \) \( ** p < 0.01, \) \( *** p < 0.001 \), interaction partner (i.p.), object (o.), else (e.). su = sum = summary, in = intro = introduction

Figure 6. This graph shows the total length of eye-gaze bouts to the interaction partner, the object and somewhere else (y-axis) over time: all seven action parts are displayed (x-axis) for ACI (left), AAI (middle) and ARI (right) condition.

hand movements which are clearly distinguishable from those observable in adult-adult interactions and which are even more accentuated than the hand movements in child-directed tutoring. Thus, ostensive stimuli are present in robot tutoring. These however change over time as we have seen: range of motion decreases drastically, whereas velocity increases slowly. We therefore hypothesize that the reason for this lies in the behavior of the learner which shapes the behavior of the tutor as stated for eye gaze behavior and hand movements by Pitsch et al. [11]. This process could be interpreted as an alignment process where the tutor starts of by clearly signaling his intention of tutoring the infant. This signal decreases during the ongoing interaction while the tutor captures the infant's attention and while observing an understanding process in the infant. The final behavior may thus be described as consisting of fragmentary cues rather than the complete and exaggerated signal. On the other hand, our results reveal that in order to create a contingent interaction with the partner, the learner needs to produce a suitable feedback. This means that although the tutor’s hand movements in robot-directed tutoring seem to be even slower and less round than in child-directed tutoring, the tutor’s eyegazing behavior in robot-directed tutoring is suggestive of a lack of appropriate social signals on
the recipient’s side: The percentage of time the interaction partner is viewed by the tutor is much lower in ARI than in ACI. The ostensive signals considered here appear practical for the robot to detect situations in which it is being tutored, but we argue that a robot cannot make use of an important ostensive stimulus such as contingency without providing the “right” signals for the interactional construct. In detail, we find that already from the introduction on: the eye-gaze behavior in the ARI situation is rather similar to that of the AAI situation, with less time of the eye-gaze being spent on the interaction partner. This is congruent with previous findings from [14]. If we hypothesize that eye-gaze is also being used in order to check for understanding of the partner, the eye-gaze behavior directly after the end of a sub-action becomes relevant. Indeed, we can see that the eye-gaze lengths in both pauses p1 and p2 are significantly longer in ACI as opposed to AAI. Thus, the parents appear to look for understanding in their infants. Interestingly, the behavior in ARI tends to be similar to the one in AAI indicating that adults behave differently towards robots. However, in p1 we see a trend for the eye-gaze lengths to be significantly longer in ARI as opposed to AAI. This might indicate that the subjects are looking out for signs of understanding in the robot as well. Yet, this behavior dramatically changes in p2 where the eye-gaze length is again decreased to the level of AAI, whereas it is even slightly increased in ACI. This may be interpreted as a reaction to missing signals of understanding from the robot. In the summary part of the action (sum), finally, the overall eye-gaze length towards the robot becomes significantly shorter than in both, ACI and AAI. In order to confirm these results and our interpretation we are planning to carry out analyses of the joint eye-gaze behavior. We hypothesize that the robot is not able to establish mutual gaze especially in the pauses which then leads to the increase of eye-gaze towards the robot.

6. Outlook

These findings suggest that ostensive signals are present in human-robot tutoring situations and may be used for the robot to learn. However, in order for the robot to elicit a contingent interaction, it needs to provide ostensive signals that indicate its understanding. Based on our observations of the infants’ behavior, these ostensive signals have to pertain to attention. That is, the robot has to provide eye gaze that signals attention and establishes joint attention as well as shared attention. Another behavior of the infants that was not modeled in the ARI condition was their attempts to reach and grasp the demonstrated objects. Further analyses need to be carried out in order to reveal the pattern of these reaching gestures - first impressions of the data suggest that they are far from random but only appear at the end of the demonstrated actions. If this is true, the reaching gestures could be interpreted as a signal that the infant has understood the goal of the action, or at least, the end of the action. Further signals which can be observed from the infants are facial expressions. Again, systematic analyses need to be carried out, but first impressions suggest that emotional feedback indicates affective reactions to the objects themselves, but also to the attention grabbing behavior of the tutor, and the reaching of the goal.

References

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

Karola Pitsch1,2, Anna-Lisa Vollmer2,1, Jannik Fritsch2, Britta Wrede1,2, Katharina Rohlfing1,2, Gerhard Sagerer1

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

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 capabilities 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 homogenous 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.

01 VP: KUCK mal; ERST nehmen wir den GRÜ: (1,0)

VP-act: g grab g lift hold g place

RC-gaz: @Ø @g >>> @g

*1 *2 *3
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 separate 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 known 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-orientates to the floor – just when VP grabs the green cup (01).

Again, VP verbally calls the child (“HANnah”), the child then re-orientates 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).

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 – re-orientates 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 re-orientates 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 proceeds to grab the red cup and move it into the blue one with another flat motion. The child disattends immediately (img.5 and 6).

This conduct confirms our hypothesis and suggests that presentational actions without modification do not seem to be appropriate for helping the child to attend 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 quantitative 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 visualizations (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.
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


Which ‘Motionese’ Parameters Change With Children’s Age?

Anna-Lisa Vollmer, Katrin Solveig Lohan, Jannik Fritsch, Katharina Rohlfing, and Britta Wrede

CoR-Lab, Bielefeld University, Germany
EU-Project iTALK
Honda Research Institute Europe

Motivation and Abstract

While it is already known that parents modify their demonstrations towards children (Brand et al., 2002; Brand et al., 2007) and that young infants aged 6 to 8 months prefer ‘motionese’ (Brand & Shallcross, 2007), little is known about whether the modified behavior can also be found in interaction with older children. Here, we therefore seek to investigate the effects of children’s age on motionese, defined as modified action demonstration (Brand et al., 2002, Rohlfing et al., 2006). In our study, parents demonstrated a function of an object (stacking cups) towards their infant and towards another adult. We analyzed parental behavior in three different age groups: parents of prelexical (8 – 11), early lexical (12 – 24) and advanced lexical (25 – 30 months olds) children. In our analysis, we use objective measurements of hand trajectories providing data about their shape and time structure. Results suggest that actions chosen to attract attention by providing more range can primarily be found in interaction with younger infants, whose attention needs more guidance. Interactions with older children seem to benefit either from the increase of children’s attention abilities or that parents use other means (such as language) to attract their attention. In contrast, parameters that appear to be more in charge of structuring the action by organizing it in motion pause seem to persist over the age and verbal capabilities.

Annotation and Data Analysis

Action Segmentation: The action of the stacking-cups and additionally, the sub-actions (a1–a3) of grasping one cup until releasing it into the end position (Fig. 1), were marked in the video. We defined:

1) the action as the whole process of transporting all objects to their goal positions;
2) a subaction as the process of transporting one object to its goal position;
3) movements as phases where the velocity of the hand is above a certain threshold; all other phases are defined as pauses.

Action Range: covered motion path divided by the distance between subaction on- and offset.

Action Pace: calculated for each movement by dividing its duration (in ms) by the duration of its preceding pause (in ms).

Total length of motion pauses: as the percentage of time of the action without movement.

Total length of eye-gaze bouts: to interaction partner as the percentage of time of the action.

Results and Discussion

Results

A repeated measures ANOVA with interaction condition (AC/AA) as within-subjects and infants’ age as between-subjects factor revealed a significant main effect for the interaction condition for all measures (p<0.001), except range. Subsequently, paired t-tests were conducted for the three age groups separately. For the range measure, we found significant differences between the conditions only in group 1 for subaction 3 (k7=2.55*) and marginal significance for subaction 2 (k7=2.15†). This suggests that the modified range of hand movements is present only in demonstrations towards pre-lexical infants. We think the reason is younger infants’ need of gestures to attract their attention. The pace measure shows significance for groups 1 (k7=4.95*) and 3 (k9=2.82†), which suggests that pace in interactions with infants of all three age groups remains higher than in the AA condition. For motion pauses, we found significant differences for age groups 2 (k10=2.79*) and 3 (k9=4.55***) and a trend for group 1 (k7=3.24†). Pauses structuring the shown action seem to be used over all age groups. For the eye gaze measure, a decrease in significance could have found between the children’s age: In the AC condition, the interaction partner was gazed at significantly longer in groups 1 (k7=3.96*†), 2 (k10=2.16*†) and 3 (k9=2.34†) and objects were gazed at significantly less in groups 1 (k7=3.98**) and 2 (k10=3.62**) suggesting that the young infants’ attention is more often checked on.

Discussion

Actions chosen to attract attention can primarily be found in interaction with younger infants, whose attention needs more guidance. Interactions with older children seem to differ due to either the increase of children’s attention abilities or that parents use other means to attract their attention. In contrast, parameters that appear to be more in charge of structuring the action seem to persist over the children’s age and their verbal capabilities.

Acknowledgment

Anna-Lisa Vollmer gratefully acknowledges the financial support from Honda Research Institute Europe for the project “Acquiring and Utilizing Correlation Patterns across Multiple Input Modalities for Developmental Learning”. Katrin Lohan gratefully acknowledges the financial support from the FP7 European Project iTALK (ICT-214668).

References

Abstract—In developmental research, tutoring behavior has been identified as scaffolding infants’ learning processes. It has been defined in terms of child-directed speech (Motherese), child-directed motion (Motionese), and contingency. In the field of developmental robotics, research often assumes that in human-robot interaction (HRI), robots are treated similar to infants, because their immature cognitive capabilities benefit from this behavior. However, according to our knowledge, it has barely been studied whether this is true and how exactly humans alter their behavior towards a robotic interaction partner. In this paper, we present results concerning the acceptance of a robotic agent in a social learning scenario obtained via comparison to adults and 8–11 months old infants in equal conditions. These results constitute an important empirical basis for making use of tutoring behavior in social robotics. In our study, we performed a detailed multimodal analysis of HRI in a tutoring situation using the example of a robot simulation equipped with a bottom-up saliency-based attention model [1]. Our results reveal significant differences in hand movement velocity, range of motion, and eye gaze suggesting that for example adults decrease their hand movement velocity in an Adult-Child Interaction (ACI), opposed to an Adult-Adult Interaction (AAI) and this decrease is even higher in the Adult-Robot Interaction (ARI). We also found important differences between ACI and ARI in how the behavior is modified over time as the interaction unfolds. These findings indicate the necessity of integrating top-down feedback structures into a bottom-up system for robots to be fully accepted as interaction partners.

I. INTRODUCTION

Learning in human children is not only a concern of an individual. It has been shown that it is a social endeavor and children get support from the social partner on multimodal levels: Adults can not only adjust their speech [2], but also their gesture [3] and motion [4], [5]. It has also been shown that children not only prefer [6], but also can benefit from these modifications [7]. This benefit has attracted attention of research in developmental robotics. The objective is here that if the interaction between a robot and its user could be designed based on the child-adult interaction, the robot – similar to the child – could obtain the more structured and enriched input and benefit from it in its learning process [1], [8], [9]. This is particularly interesting for learning actions, since – without support and only by observation – it is difficult for a robot to decide what and when to imitate [10], [11]. With these problems in mind, it has been suggested that using modifications in tutors’ behavior, a robot could learn to detect the meaningful structure of the demonstrated action [1], [8]. However, we do not know yet the crucial characteristics that establish a natural tutoring situation. It has been assumed that a robot – because of its immature cognitive capabilities – can trigger a tutoring behavior in its interaction partner [12]. However, this assumption has barely been studied. Recently, a study by Herberg and his colleagues [13] investigated the question whether people will modify their actions for computers. They presented a picture of an interaction partner to the subjects, which varied in dependence on the condition: a child, an adult and a computer together with a monitor and a mounted camera on it in a second condition [13]. The authors found that subjects modified their actions when speaking to a computer. The modifications differed from how they interacted with a picture of a child or adult. Herberg and his colleagues [13] interpret the difference in terms of assigning – to the persons, but not to the computer – the capability of reasoning about goals. However, it is difficult to expect from a user to assign some capabilities just from viewing a picture. It has been shown that subjects, when asked to speak to an imaginary infant, were not able to produce speech that exhibits all the features that are characteristic for motherese as it is produced in real adult-infant interactions [14]. The results from Herberg et al. should thus be interpreted with caution. Also, interactions with a computer are differently processed by subjects than interactions with robots especially with respect to the assignment of intentions. In an fMRI study Krach et al. [15] have shown that the brain area that is generally associated with theory-of-mind (thus, the reasoning about the others intentions) is significantly stronger activated when the subjects thought they were interacting with a humanoid robot than when they thought they were interacting with a computer. Contingency describes situations in which two agents socially interact with each other and Csibra and Gergely showed that contingency is a characteristic aspect of social interaction [10]. In the study published by Herberg et al. there is no possible reactivity in the interaction partner, so we argue that social interaction cannot take place.

In this work we therefore present results from real interactions with an embodied simulated robot based on the assumption that real interaction is needed in order to coordinate...
the behavior with the partner and to open up for mutual influence [16]. We think that only such a scenario can create an environment in which we can find out about the crucial characteristics of a natural tutoring situation.

In our study, similar to Herberg et al. [13], we pursued the question of whether people will modify their actions when interacting with a machine. In contrast to Herberg et al., who used a computer, we investigated the interaction with a virtual robot. For our purpose, we analyzed real interactions – and not just a picture of the partner as in the previous study – with the artificial system and compared the results to the results obtained from real interactions with a child and an adult. For our analysis, we applied a battery of measurements allowing for a fine-grained analysis of performed motions and their changes in the interaction as it unfolds.

II. EXPERIMENT

Data was obtained in two experiments. The data on adult-child interaction was obtained in the Motionese experiment, which is based on the same setting as in [8] and [1]. The data on human-robot interaction was obtained in the second experiment.

A. Motionese Experiment (ME)

1) Subjects: The Motionese Corpus consists of infant- and adult-directed interactions. We selected the younger group comprising 12 families of 8 to 11 months old children. Both parents were asked to demonstrate functions of 10 different objects to their children as well as to their partners or another adult. In the following, we focus on the analysis of the stacking cups task, because it offers the best comparability in motion performance. We further selected a subgroup of 8 parents (4 fathers and 4 mothers) for the ACI and a subgroup of 12 parents (7 fathers and 5 mothers) for the AAI, because of the quality of the video, sound and due to the way in which the action was performed. More specifically, the order in which the cups of the considered stacking-cups task are put together can vary: We selected only those parents, who started the task by putting the first cup into the target cup which means putting the green cup into the blue one (see Fig. 3 a1).

2) Setting: Parents were instructed to demonstrate a stacking-cups task to an interaction partner. The interaction partner was first their infant and then an adult. Fig. 1 illustrates the top-view of the experimental setup, and shows sample image frames of cameras which were set behind the parent and the interaction partner and focused on each of them. The stacking-cups task was to sequentially pick up the green (a1), the yellow (a2), and the red (a3) cup and put them into the blue one on the white tray.

B. Robot-Directed Interaction Experiment (RDIE)

1) Subjects: 31 adults (14 females and 17 male) participated in this experiment 7 out of which were parents as well. Out of this group, we selected 12 participants (8 female and 4 male), who performed the task in a comparable manner.

2) Setting: The participants were instructed to demonstrate several objects to an interaction partner, while explaining him/her how to do it (Fig. 2). Again we chose the stacking-cups task for analysis. The interaction partner was an infant-like looking virtual robot with a saliency-based visual attention system [1]. The robot-eyes will follow the most salient point in the scene, which is computed by color, movement, and other features (see [1] and Fig. 4).

III. DATA ANALYSIS

The goal of this paper was to analyze tutoring behavior from two perspectives, Motionese and Contingency. For this reason, we analyzed Motionese and Contingency features. We coded the videos semi-automatically to obtain data for the 2D hand trajectories and the eye gaze directions.

A. Annotations

For all annotations, we used the video captured by camera (cam) 1, see Fig. 1 and 2. It shows the front view on the demonstrator and is therefore best suited for action, movement, and gaze annotations, which are discussed in detail below.

1) Motionese: Action Segmentation: For analyzing the data, the action of the stacking-cups and additionally, the sub-actions (a1-a3) of grasping one cup until releasing it into the end position (Fig. 3) were marked in the video. We defined

1) action as the whole process of transporting all objects to their goal positions.
2) subaction as the process of transporting one object to its goal position.
movement as phases where the velocity of the hand is above a certain threshold. All other phases are defined as pauses.

**Hand Trajectories:** The videos of the two experiments were analyzed via a semiautomatic hand tracker system (Fig. 4). The system is written as a plugin for a graphical plugin shell, iceWing [17], and makes it possible to track both hands with an Optical Flow based algorithm, Lucas & Kanade [18]. The system allows manual adjustment in case of tracking deviation. We used this tracking system instead of the previously used 3D body model system, [8], since 3D results in [8] were not significant, we focused on 2D analyses which provide to show more stable results. Additionally, the new system is easily accessible for non-expert users.

![Fig. 4. In the left picture, the red and violet circles depict the tracking regions which are tracked by the hand tracker system. The points in the middle of the circles are the resulting points for the 2D hand trajectory. In the right picture, the virtual robot we used is shown.](image)

**2) Contingency:**

**Eye Gaze:** In annotating the eye gaze directions with the program Interact (Mangold), we distinguished between looking at the interaction partner and looking at the object (Fig. 5).

![Fig. 5. These three pictures show the difference between looking to the object (left), looking to the interaction partner (middle) and looking somewhere else (right).](image)

**B. Measures**

For quantifying Motionese and Contingency, we computed seventeen variables related to the 2D hand trajectories derived from the videos and the eye gaze bout annotations produced with Interact.

1) **Motionese:** We operationalized Motionese in terms of velocity, acceleration, pace, roundness, and motion pauses as defined in [8]. Rohlfing et al. automatically segmented the task into movements and pauses based on hand velocity.

**Velocity** was computed using the derivative of the 2-dimensional hand coordinates of the hand which performed the action per frame. Rohlfing et al. did not find a significant effect for velocity for the 3D posture tracking data. Their 2D hand tracking data showed the statistically significant trend that hand movement in AAI is faster than in ACI.

**Acceleration** is thus defined as the second derivative of the hand trajectory.

**Pace** was defined for each movement by dividing the duration of the movement (in ms) by the duration of the preceding pause (in ms). For pace, Rohlfing et al. found nearly significant differences comparing ACI and AAI. Their results suggest that pace values in ACI are lower than in AAI.

**Roundness** of a movement was defined by covered motion path (in meters) divided by the distance between motion on- and offset (in meters). Thus, a higher value in roundness means rounder movements. Rohlfing et al. found that hand movement is significantly rounder in AAI compared to ACI.

**Frequency of motion pauses** was defined as the number of motion pauses per minute. Therefore, the number of motion pauses was computed automatically using the above-mentioned segmentation, see Fig. 3. Further, the **average length of motion pauses** (in frames) and **total length of motion pauses** as the percentage of time of the action without movement were computed.

Additionally, we focused on the trajectory during the actual transportation of the cups, when performing the task. For each video and setting, the exact video frames of the beginnings and ends of the transportation for each of the three cups were annotated by hand, again see Fig. 3. This way, we were able to define variables for each individual subaction (a1, a2, a3) and also detect changes in the demonstrator’s behavior in the course of fulfilling the task.

**Subaction specific velocity** was computed as the average velocity for subactions a1, a2, and a3 each.

**Subaction specific acceleration** was computed analogously as the average acceleration for subactions a1, a2, and a3.

**Range** was defined as the covered motion path divided by the distance between motion, i.e. subaction, on- and offset.

**Action length** denoted the overall action length and was measured from the beginning of subaction a1 to the end of subaction a3.

2) **Contingency:** J.S. Watson thinks of contingency as the human infant’s means for detecting socially responsive agents and therefore postulates the existence of an innate contingency detection module as one of the most fundamental innate modules. He formally defines the contingent temporal relation of two events, for example a response R and a stimulus reward S, as two conditional probabilities. The first, called the sufficiency index, measures the probability of a stimulus reward S given a span of time t following a response R, \( P(S*|Rt) \). The second, called the necessity index, measures the probability of the response given time span t prior to the reward stimulus, \( P(R|tS*) \) [19]. "Contingency detection is crucially involved in an infant’s progressively developing awareness of his or her internal affective states" [10]. "The discovery that another agent’s gaze is a cue worthy of monitoring relies on the infant’s ability to detect the contingency structure in interactions with that agent" [20]. The Contingency of the interactions was quantified in terms of variables related to eye gaze, as defined in [21] for measuring interactiveness.

**Frequency of eye-gaze bouts to interaction partner,** i.e. eye gaze bouts per minute, was computed from the Interact annotations. Also, the **average length of eye-gaze bout to**
interaction partner and the total length of eye-gaze bouts to interaction partner as the percentage of time of the action spent gazing at the interaction partner were computed. Brand et al. found that infants received significantly more eye-gaze bouts per minute [21], so the frequency of eye-gaze bouts to the interaction partner was significantly higher in ACI than in AAI. The total and average length of eye-gaze bouts to the interaction partner in their study was significantly greater in ACI than in AAI. Equivalent measures were calculated for the eye gaze on the demonstrated object. Namely, we obtained values for frequency of eye-gaze bouts to object, average length of eye-gaze bout to object, and total length of eye-gaze bouts to object as the percentage of time of the action spent gazing at the object.

IV. RESULTS

Table I depicts the results of the study.

A. Motionese

A non-parametric test (Mann-Whitney U test) was run for all pairs of samples, ACI vs. AAI, ACI vs. ARI, and AAI vs. ARI. For velocity, the test revealed significant differences for ACI vs. AAI and ACI vs. ARI, and highly significant differences when testing AAI vs. ARI. These results show that in AAI hand movements are significantly slower than in ACI and hand movements in ACI are significantly slower than in AAI.

For the subaction specific velocity measure, which only takes into account the hand movement during the transportation of the respective cup, the results were even more significant. For all pairs of conditions, we also found significant differences for all three subactions. These results clearly show that in AAI hand movements are very fast compared to ACI and ARI and additionally that hand movement is slowest in the ARI condition. Also note that for all conditions the mean values increase for the consecutive subactions. This also holds for the variances, i.e. mean and variance for the velocity of hand movement in subaction a3 are greatest. In the ARI, the rate in which the mean values increase is slowest.

The tests showed no significance for acceleration in ACI vs. AAI and ACI vs. ARI, but show a trend which is that acceleration of hand movement in ACI is slower than in AAI and greater than in ARI. They show significant results for AAI vs. ARI conditions, i.e. in AAI, hand movement acceleration is significantly greater than in the ARI.

Viewing this measure again for only the transportation of the cups in the different subactions, the test results reveal significant differences and statistical trends for all pairs of conditions and almost all subactions. Results suggest that subaction specific acceleration of hand movement is lower in ACI than in AAI. The mean values for each consecutive subaction increase for both conditions, so that results for a2 revealed significance, whereas results for a1 and a3 show a trend. Also hand movement acceleration is highly significantly lower in ARI than in AAI. For ACI vs. ARI results reveal significance for a3 and a trend for a2. Note again that for ARI mean values increase at a lower rate.

Pace results revealed highly significant differences for AAI vs. ARI and significant differences for ACI vs. ARI and ACI vs. AAI. The latter confirms the findings in [8] that pace in AAI is higher than in ACI. The results indicate ARI having significantly slower pace than AAI and ACI and AAI having significantly slower pace than AAI. Note that the variance of pace in ARI is very small.

The results for the roundness measure show that movement is roughest in AAI compared to the other two conditions. Differences between ACI and AAI, and AAI and ARI are significant. No significance was found for ACI vs. ARI. The range measure suggests that ARI exhibits the greatest range and for this reason most exaggerated movement for all subactions a1 to a3 and also that range is greater in ACI than in AAI. For ACI vs. AAI results revealed significance for subactions a2 and a3, and a trend for a1. For ACI vs. ARI solely results for subaction a1 showed significance, a2 and a3 did not. For AAI vs. ARI subactions a1 to a3 revealed significance.

When analyzing motion pauses, tests revealed that in AAI the frequency of motion pauses is significantly lower than in ACI and ARI. For ACI vs. ARI no significant differences were found.

The average length of motion pauses is significantly smaller in the AAI condition than in the ACI and the ARI condition. For ACI vs. ARI test results did not show significance, but a statistical trend which is that values for ARI are greater than for ACI. Comparing the total length of motion pauses, results are again significant for ACI vs. AAI and AAI vs. ARI. Hence, results show that the total length of motion pauses is significantly smaller in AAI than in ACI and ARI.

The overall action length is greater in ARI than in ACI, where the action length is again greater than in AAI. Adults thus take more time, when demonstrating object functions to children compared to demonstrating them to adults, but they take even more time when demonstrating objects to a robot. The tests showed that differences between ACI and AAI are significant, as well as differences between AAI and ARI. Differences between ACI and ARI were marginally not significant. Thus, in general the movement in ARI appears to be even more accentuated than in ACI.

B. Contingency

Most interestingly the results for eye gaze show a completely different picture. The contingency measures revealed for frequency of eye-gaze bouts to interaction partner significant differences for ACI vs. AAI and ACI vs. ARI, but not for AAI vs. ARI. In ACI eye-gaze bouts to the interaction partner were most frequent.

Testing the average length of eye gaze bout to interaction partner, we found on average significantly longer bouts in ACI than in AAI and ARI and a trend for AAI vs. ARI. For total length of eye-gaze bouts to interaction partner they showed that in ACI significantly more time was spent gazing at the interaction partner than for AAI and ARI. Differences
between AAI and ARI again are not significant.

For eye-gaze to the object, we found that frequency of eye-gaze bouts to object is significantly lower in ARI than in the other two conditions, ACI and AAI. Differences in ACI and AAI were not significant.

Average length of eye-gaze bout to object was significantly smaller for ACI than for AAI and ARI. Here, differences between AAI and ARI were not significant.

The same is true for the measure total length of eye-gaze bouts to object. Values are significantly lower in ACI than in AAI and ARI, where again differences between AAI and ARI did not exhibit significance.

Fig. 6. This graph shows the mean frequency of eye-gaze bouts to interaction partner and object (y-axis) over the whole action in every condition (x-axis).

V. DISCUSSION AND CONCLUSION

In sum, our results show a differentiated picture for modifications in human-robot interaction. On the one hand, we have
found that a robot receives even more strongly accentuated input than an infant: almost all hand movement-related variables, when pooled over the whole action sequence, showed a significant difference, or at least a trend, between the three conditions with a clear ordering (AAI > ACI > ARI). ARI movements can thus be characterized as slower (velocity, acceleration, and pace), more exaggerated (range) than AAI, and less round (roundness) than AAI movements. In contrast to ACI, where the tutoring behavior seems to bear lots of variability, in the ARI, more stability could be observed. This suggests that ARI allows to control for the parameters of the learner and is thus a promising method for studying tutoring behavior. On the other hand, the contingency measurements show less contingent eye gazing behavior in ARI than in ACI (frequency and length of eye-gaze bouts to interaction partner). These results raise an interesting question: Why is the behavior of the tutors in the ARI condition less contingent than in the ACI condition? As contingency is a bi-directional phenomenon, it is likely to be related to the robot’s feedback behavior. Indeed, while the frequency of motion pauses is similar in ARI and ACI, the length of motion pauses is significantly longer in ARI than in AAI and ACI indicating that the tutor is waiting possibly in vain - for a sign of understanding from the robot. The lower amount of eye-gaze bouts to the interaction partner in ARI as opposed to ACI could be interpreted similarly: as the tutor does not receive the expected feedback of understanding from the robot, s/he does not search for eye-contact with the robot. In future research, we will focus more closely on feedback behavior and identify the important signals in a bi-directional interaction.

These results have important consequences for human-robot interaction in developmental robotics. They indicate that the behavior of the robot shapes the behavior of the tutor. Although all tutors showed strong modifications in their movement behavior towards a robot, thus stressing important aspects of the demonstrated action, they did not increase their contingency behavior as other tutors would do in interactions with infants. Even though the purely reactive behavior of the robot in our study does induce parent-like teaching (as indicated in a qualitative study by Nagai et al. [22]), it does not seem to be sufficient to produce a contingent interaction. As studies show, contingent behavior is an important feature for learning in human development. Thus, in order for robots to be able to learn from a human tutor, they should have the capability to engage in a contingent interaction. Further analyses need to be carried out with the goal to reveal what exactly causes the tutor to decrease her contingent behavior in ARI.

A. Summary

For a short summary of our results see Table II.

ACKNOWLEDGMENT

Anna-Lisa Vollmer gratefully acknowledges the financial support from Honda Research Institute Europe for the project ‘Acquiring and Utilizing Correlation Patterns across Multiple Input Modalities for Developmental Learning’. Kerstin Fischer, Katrin Lohan, Karola Pitsch, Katharina Rohlfing, and Britta Wrede gratefully acknowledge the financial support from the FP7 European Project ITALK (ICT-214668). Yukie Nagai gratefully acknowledges the financial support from Honda Research Institute Europe for the project ‘Designing Human-Robot Interaction based on/toward Understanding Parent-Infant Interaction’.

REFERENCES