People Modify Their Tutoring Behavior in Robot-Directed Interaction for Action Learning

Anna-Lisa Vollmer*, Katrin Solveig Lohan[x], Kerstin Fischer‡, Yukie Nagai‖, Karola Pitsch‖, Jannik Fritsch§,
Katharina J. Rohlfing‖, and Britta Wrede‖
*[x]‖Bielefeld University, CoR-Lab, Applied Informatics Group, Bielefeld, Germany, http://www.cor-lab.de
*Email: avollmer@cor-lab.uni-bielefeld.de
‡Institute of Business Communication and Information Science, Sonderborg, Denmark
§Honda Research Institute Europe GmbH, Offenbach, Germany

***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, 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 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 reactiveness 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.

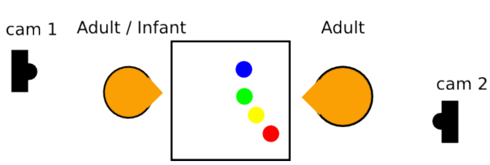

Fig. 1. Motionese Setting, there are two cameras which are recording the scene. The interaction partners are seated across from each other and the object is laid on the table in front of the tutor.

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).

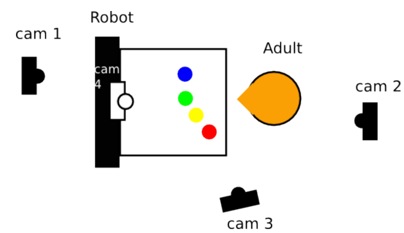

Fig. 2. Robot-directed Interaction Setting, there are four cameras which are recording the scene. The subject is seated across from the robot and the object is laid on the table in front of the tutor.

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.

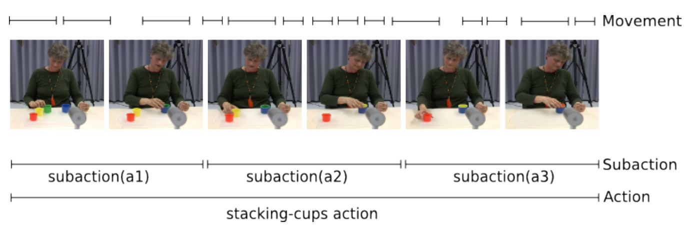

Fig. 3. This graphic shows an example for the structure of an 'Action', 'Subaction', and 'Movement'.

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.

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 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.

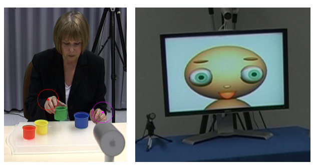

Fig. 4. In the left picture, the red and violet circles depict the tracking regions which are tracked by the hand tracker system. The points in the middle of the circles are the resulting points for the 2D hand trajectory. In the right picture, the virtual robot we used is shown.

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).

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 ARI 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 smaller 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 ACI 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 roundest 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

Variable	ACI		ARI		AAI		ACI vs AAI	ACI vs ARI	AAI vs ARI
	M	*SD*	*M*	*SD*	*M*	*SD*	*Z*	*Z*	*Z*
velocity	0.17	0.06	0.29	0.07	0.12	0.03	−3.086**	−2.315*	−3.926***
velocity a1	4.33	1.71	7.89	2.01	2.95	0.82	−2.855**	−2.006*	−4.041***
velocity a2	5.9	2.25	11.14	2.38	3.59	1.16	−3.318***	−2.546*	−4.157***
velocity a3	7.24	2.42	13.93	3.75	4.83	1.66	−3.163**	−2.469*	−3.984***
acceleration	0.05	0.03	0.08	0.03	0.03	0.01	−1.697+	−1.929+	−3.637***
acceleration a1	1.18	0.64	1.75	0.56	0.78	0.37	−1.929+	−1.543	−3.233***
acceleration a2	1.58	1.06	2.93	0.84	0.84	0.34	−2.700**	−1.852+	−4.157***
acceleration a3	2.67	1.27	3.88	1.53	1.19	0.57	−1.929+	−2.777**	−3.926***
pace	17.68	32.78	56.03	39.69	4.25	1.98	−2.415**	−1.774*	−3.703***
roundness	2.87	2.49	7.26	2.71	1.73	0.30	−2.855***	−0.231	−4.099***
total length m.p.	16.89	11.29	1.46	2.9	28.08	11.25	−3.091**	−1.543	−4.270***
frequency m.p.	37.88	14.28	23.28	10.56	40.05	7.1	−2.006*	−0.154	−3.175***
average length m.p.	5.92	3.68	0.58	1.14	11	5.39	−3.174**	−1.852+	−4.270***
range a1	2.54	1.07	1.76	0.42	4.09	1.52	−1.929+	−2.392*	−3.926***
range a2	1.69	0.41	1.33	0.18	1.81	0.44	−2.083*	−0.772	−3.175***
range a3	1.45	0.25	1.24	0.18	1.64	0.4	−2.392*	−1.312	−3.002**
action length	9.68	4.2	3.65	1.11	14.41	5.66	−3.240***	−1.697+	−4.157***
total length eye-gaze to i.p.	36.38	22.61	7.78	9.65	9.99	13.25	−2.815**	−2.633**	−0.539
frequency eye-gaze to i.p.	33.96	10.13	11.84	14.43	8.93	8.11	−2.893**	−3.640***	−0.120
average length eye-gaze to i.p.	0.94	0.39	0.22	0.29	0.45	0.41	−3.127***	−2.556**	−1.438+
total length eye-gaze to o.	59.48	23.17	90.74	11.31	88.87	14.13	−2.971**	−2.788**	−0.360
frequency eye-gaze to o.	35.34	6.43	28.12	14.73	12.68	5.83	−1.852+	−3.626***	−3.233***
average length eye-gaze to o.	1.3	0.76	5.74	3.36	10.04	9.72	−3.086**	−3.549***	−0.924

TABLE I
RESULTS OF MEAN, STANDARD DEVIATION, MANN-WHITNEY U TEST, $+p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, MOTION PAUSES ($m.p.$), INTERACTION PARTNER ($i.p.$), OBJECT ($o.$). (DUE TO RESULTS IN [8] AND [21], WE PERFORMED A ONE-TAILED ANALYSIS FOR PACE, ROUNDNESS, AND AVERAGE LENGTH EYE-GAZE TO $i.p.$)

Compared to AAI, ACI shows	Compared to ACI, ARI shows	Compared to AAI, ARI shows
slower hand movement	slower hand movement	slower hand movement
lower hand movement acceleration	lower hand movement acceleration	lower hand movement acceleration
smaller pace	smaller pace	smaller pace
less round movement		less round movement
greater range and therewith more exaggerated movement	greater range and therewith more exaggerated movement in the first subaction	greater range and therewith more exaggerated movement
higher frequency of motion pauses		higher frequency of motion pauses
greater average length of motion pauses	greater average length of motion pauses	greater average length of motion pauses
greater total length of motion pauses	greater total length of motion pauses	greater total length of motion pauses
longer action	longer action	longer action
more frequent eye-gaze bouts to the interaction partner	less frequent eye-gaze bouts to the interaction partner	
on average longer eye-gaze bouts to the interaction partner	on average shorter eye-gaze bouts to the interaction partner	
more time spent gazing at the interaction partner	less time spent gazing at the interaction partner	
higher frequency of eye-gaze bouts to object	lower frequency of eye-gaze bouts to object	lower frequency of eye-gaze bouts to object
smaller average length of eye-gaze bout to object	greater average length of eye-gaze bout to object	
smaller total length of eye-gaze bouts to object	greater total length of eye-gaze bouts to object	

TABLE II
THIS TABLE SHOWS A SHORT SUMMARY OF OUR RESULTS.

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.

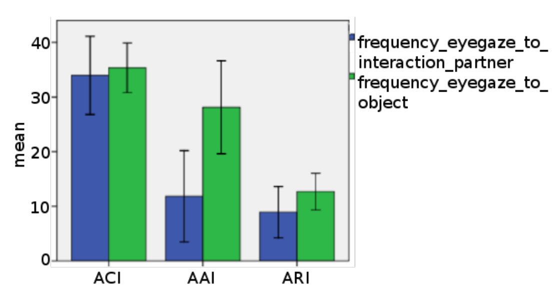

Fig. 6. This graph shows the mean frequency of eye-gaze bouts to interaction partner and object (y-axis) over the whole action in every condition (x-axis).

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. [12]), 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

[1] Y. Nagai and K. Rohlfing, "Can motionese tell infants and robots what to imitate?," in *Proceedings of the 4th International Symposium on Imitation in Animals and Artifacts*, 2007, pp. 299–306.
[2] A. Fernald and C. Mazzie, "Prosody and focus in speech to infants and adults." *Developmental Psychology*, vol. 27, no. 2, pp. 209–21, 1991.
[3] J. Iverson, O. Capirci, E. Longobardi, and M. Cristina Caselli, "Gesturing in mother-child interactions," *Cognitive Development*, vol. 14, no. 1, pp. 57–75, 1999.
[4] L. Gogate, L. Bahrick, and J. Watson, "A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures," *Child Development*, vol. 71, no. 4, pp. 878–894, 2000.
[5] R. Brand, D. Baldwin, and L. Ashburn, "Evidence for'motionese': modifications in mothers' infant-directed action," *Developmental Science*, vol. 5, no. 1, pp. 72–83, 2002.
[6] R. Brand and W. Shallcross, "Infants prefer motionese to adult-directed action," *Developmental Science*, vol. 11, no. 6, pp. 853–861, 2008.
[7] N. Masataka, *The Onset of Language*. Cambridge University Press, 2003.
[8] K. Rohlfing, J. Fritsch, B. Wrede, and T. Jungmann, "How can multimodal cues from child-directed interaction reduce learning complexity in robots?" *Advanced Robotics*, vol. 20, no. 10, pp. 1183–1199, 2006.
[9] B. Wrede, K. Rohlfing, M. Hanheide, and G. Sagerer, "Towards learning by interacting."
[10] G. Csibra and G. Gergely, "Social learning and social cognition: The case for pedagogy," *Processes of change in brain and cognitive development. Attention and performance*, vol. 21, 2005.
[11] M. Carpenter, J. Call, and M. Tomasello, "Twelve-and 18-month-olds copy actions in terms of goals," *Developmental Science*, vol. 8, no. 1, pp. 13–20, 2005.
[12] Y. Nagai, C. Muhl, and K. Rohlfing, "Toward designing a robot that learns actions from parental demonstrations," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 3545–3550.
[13] J. Herberg, "Audience-contingent variation in action demonstrations for humans and computers," *Cognitive Science: A Multidisciplinary Journal*, vol. 32, no. 6, pp. 1003–1020, 2008.
[14] M. Knoll and L. Scharrer, ""acoustic and affective comparisons of natural and imaginary infant-, foreigner- and adult-directed speech," 2007, pp. 1414–1417.
[15] S. Krach, F. Hegel, B. Wrede, G. Sagerer, and T. Binkofski, F.;Kircher, "Can machines think? direct interaction and perspective taking with robots investigated via fmri," *PLoS ONE*, vol. 3, 07/2008 2008. [Online]. Available: http://www.plosone.org/article/info597
[16] A. Fogel and A. Garvey, "Alive communication," *Infant Behavior and Development*, vol. 30, no. 2, pp. 251–257, 2007.
[17] F. Loemker, 2007, `http://icewing.sourceforge.net`.
[18] `http://opencv.willowgarage.com/wiki/CvReference`.
[19] J. Watson, "Contingency perception in early social development," *Social perception in infants*, pp. 157–176, 1985.
[20] I. Fasel, G. Deak, J. Triesch, and J. Movellan, "Combining embodied models and empirical research for understanding the development of shared attention," in *Proceedings of the 2nd International Conference on Development and Learning*, 2002, pp. 21–27.
[21] R. Brand, W. Shallcross, M. Sabatos, and K. Massie, "Fine-grained analysis of motionese: Eye gaze, object exchanges, and action units in infant-versus adult-directed action," *INFANCY*, vol. 11, no. 2, pp. 203–214, 2007.