Contingency allows the robot to spot the tutor and to learn from interaction

Katrin S. Lohan[||]*, Karola Pitsch[||], Katharina J. Rohlfing[||][x], Kerstin Fischer[§], Joe Saunders[‡], Hagen Lehmann[‡], Chrystopher Nehaniv[‡] and Britta Wrede[||]

[§]Institute of Business Communication and Information Science, Sonderborg, Denmark

[‡]University of Hertfordshire, Adaptive Systems Research Group

[||]Bielefeld University, CoR-Lab, Applied Informatics Group

*Email: klohan@cor-lab.uni-bielefeld.de

[x]Emergentist Semantics

***Abstract*—Having the vision of an artificial system learning from a human tutor, our aim is to improve the social interaction between a robot and its user in tutoring scenarios. For this aim, we first introduce a contingency module that is developed to elicit tutoring behavior, which we then evaluated by implementing this module on the robotic platform iCUB and within an interaction with the users. For the evaluation of our system, we considered not only the participant's behavior but also the system's logfiles as dependent variables (as it was suggested in [12] for the improvement of HRI design. We further applied Sequential Analysis as a qualitative method that provides micro-analytical insights into the sequential structure of the interaction. This way, we are able to investigate a closer interrelationship between robot's and tutor's action and how they respond to each other. We focus on two cases: In the first case, the system module was reacting to the interaction partner appropriately; in the second case, the contingency module failed to spot the tutor. We found that the contingency module enables the robot to engage in an interaction with the human tutor, who orients to the robot's conduct as appropriate and responsive. In contrast, when the robot did not engage in an appropriate responsive interaction, the tutor oriented more to the object while gazing less at the robot.**

I. Introduction

From learning by observation, robotic research has moved towards investigations of learning by interaction. This research is inspired by findings from developmental studies on human children and primates pointing to the fact that learning takes place in a social environment. Accordingly, instead of just responding to and memorizing a signal, a learner receives support from the social partners, the interaction with them, the created situation and her or his own experience about such interactions (Cangelosi et al., 2009 [2]). Recently, driven by the idea that learning through observation or imitation is limited because the observed action not always reveals its meaning, scaffolding or bootstrapping processes supporting learning received increased attention. It is studied how a learner is actually provided with additional social information that is provided by a teacher or a tutor who demonstrates what is crucial to pay attention to, e.g. the goal, the means, the constraints of a task (Zukow-Goldring, 2006 [24]). In these processes, it is essential that the tutor makes sure that the learner is receptive and ready to learn. The reciprocal contribution, i.e. the guidance of attention by tutor on the one hand and the manifestation of receptivity by a learner on the other hand, seems to follow certain interactive regularities (Clark, 1992; Fogel & Garvey, 2007 [8]; Pitsch et al. 2009 [17]). The function of these regularities has been investigated in approaches towards natural pedagogy (Csibra & Gergely, 2006 [3]; 2009 [4]). More specifically, Senju and Csibra (2008 [20]) have shown that children follow a social information conveyed by the direction of the eye gaze (i.e. they look where somebody else is looking) more reliably when both, eye-contact and motherese (child-directed speech) in addressing the child verbally proceeds the social information. This way, the social information seems to be framed in ostensive cues that also provide a sequential organization of the information conveyed: The tutor is addressing the child and the child feedbacks her or his attention focus (Pitsch et al. [17]; Estigarribia & Clark, 2007 [6]; Fogel & Garvey, 2007 [8]).

For robotic research that takes its inspiration from developmental approaches, it is essential to penetrate the concrete mechanisms of such reciprocal contribution. The motivation is that once a system is equipped with mechanisms that make it sensitive to the signals of the tutor and feedbacks its attention focus, an advantage of this social interaction can be taken and a system can learn within this interaction.

II. Model of Contingency detection

Watson [22] describes, *contingency* as a relation between a behavior and a subsequent stimulus occurring between two interaction partners serving as a powerful social signal. The detection of *contingency*, thus can be viewed as a quantitative measure providing hints about the involvement of the interaction partners and the acceptance of a robot as a social learner [11]. In a natural interaction loop, the interlocutors maintain the contingency by their turn taking behavior and their means of joint attention [17]. Monitoring mutual behavior of both interaction partners has been found crucial for detecting contingency in a tutoring situation [17]. In tutoring situations, it has been suggested that a robot can benefit from the tutor's ability to adapt to its capabilities. In order to take advantage of this adaptation, the robot needs, however, to be responsive to the tutor and, thus, to encourage her or him to interact (Pitsch et al. 2009 [17], Estigarribia and

Clark, 2007 [6]; Fogel and Garvey, 2007 [8]). Therefore, to be able to actually implement the ability of showing contingent behavior onto a robot the means for detecting ostensive cues need to be modeled. In this paper, we propose our attempt to model the ability to detect contingency by the interlocutor. Our operationalization of this ability is based on two behaviors that were observed as crucial in tutoring situations: One ostensive cue that has been found crucial for monitoring other's behavior is the gazing behavior of the interaction partners [7]. The other behavior that our contingency detector is taking into account, is a form of tutor's modifications in action performance i.e. looming action, as we will explain below in more details.
We therefore equipped our robot iCub with additional sensors that allow us to analyze the current interaction with regard to the gazing and looming behavior of the tutor and the robot.

A. Gazing behavior

For the implementation, the gazing behavior of the tutor is divided into three classes (see Fig.1). The classification is realized by a geometrical analysis if the orientation (nF) of the tutor's head is directed towards the object, the robot's face or elsewhere.

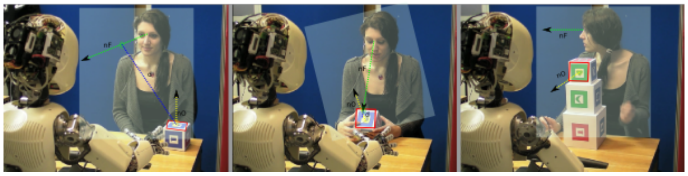

Fig. 1. Classification of the tutor's gazing behavior: a) tutor gazing towards the robot, b) tutor gazing towards the object and c) tutor gazing elsewhere

B. Looming and pointing behavior

Since child-directed action was shown to be important in a tutoring situation, Matatyaho and Gogate [14] investigated further the kind of action that is typically applied. They found that looming action, which is an action that describes a movement of a tutor moving an object towards a learner's face, is used more frequently than upward or backward motions in temporal synchrony with the spoken words. This looming motion is likely to highlight novel word-object relations [9].

According to these findings, we formulated looming behavior of the tutor - while holding an object - as a single-handed movement towards the robot and thus approach a certain distance close to the robot. In addition, if the human tutor is moving an object by hand towards the robot and reaches the Dmin (see Fig. 2) the robot responses by trying to point at the object.

C. Data Collection

The behavior of the tutor was captured by a different kind of information: We implemented and integrated some modules to detect the objects, face and whole body skeleton of the tutor. We used the FaceAPI [13] tool, to locate and tract the face of the tutor (see Fig. 3 c)). The kinect [21] 3D sensor was used to track a whole body skeleton of the tutor in order to capture the information of the head position. This was necessary because the FaceAPI is not constantly tracking the gaze; in addition it was necessary to capture the movement of the tutor's arms to calculate the looming movement of the tutor (see Fig. 3 b)). In our set-up the 3D coordinates of the objects are detected and marked with ARToolkit [10] markers (see Fig. 3 a)). To sum this up the gazing classification is based on the data of the FaceAPI, the kinect as backup and the data from the ARToolKit to locate if the tutor is looking at the object. To find out if the tutor is showing a looming behavior we used the data of the kinect and the 3D coordinates of the objects.

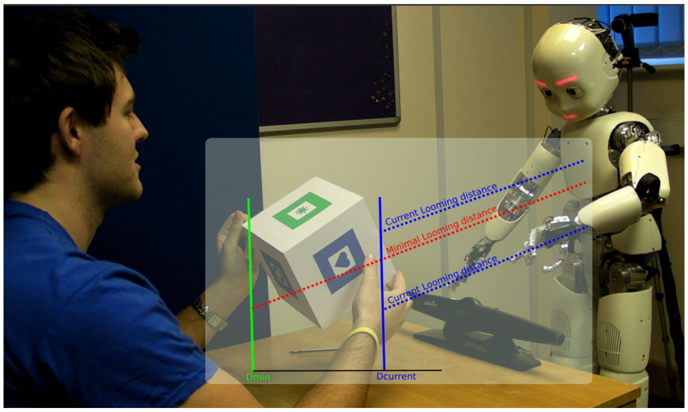

Fig. 2. Looming behavior: **Dmin** is the minimal distance that must be reached with an object and an hand of the tutor to activate the pointing behavior of the robot. **Dcurrent** represents the current distance between the hand and the object detected by the robot.

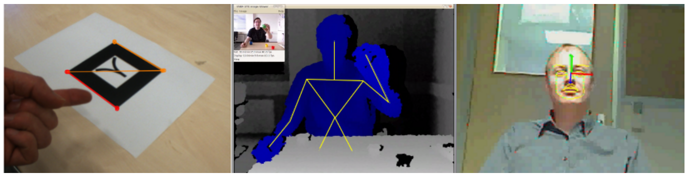

Fig. 3. a) shows an ARToolKitt [10] marker that is highlighted by the detection, b) is showing a person detected by the kinect [21] giving the skeleton of the whole body and c) is showing a person tracked by the FaceAPI [13].

D. Contingency

The structure of the robot system can be seen in Fig.4. The iCub robot is connected via YARP [15] with the system. The whole system is storing and exchanging data via an active memory based on XCF [23]. The contingency module is informed by the active memory if the tutor is gazing at the robot, at the object or somewhere else and if the tutor is presenting the object to the robot or not. Also the current behavior of the robot is known by the contingency module. When measuring contingency we take both interaction partners into account see Fig.5. Contingency is measured by the two variables necessity- and sufficieny-index.
According to Watson the necessity index describes the forward probability of a consequence given a (hypothesized) cause. From the robot's perspective, this means the probability

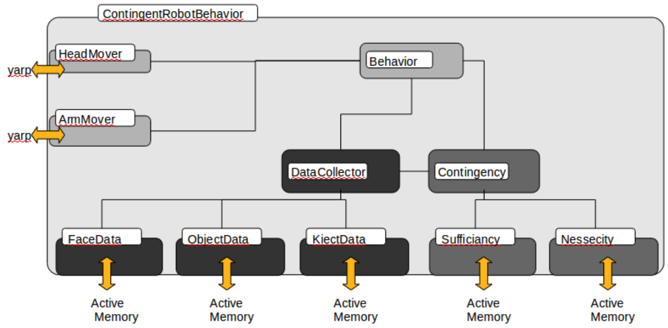

Fig. 4. The system is connected via YARP to the robot. We used the active memory to exchange and store data while running the system.

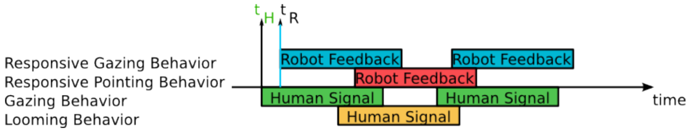

Fig. 5. At the time tH the behavior of the human tutor is detect and at the time tR the Robot is giving a responsive feedback. tH-tR is the reaction time of the system.

that the subject's gaze lies on a certain object X, given that the robot has previously been looking at X. The sufficieny index measures, if there are also other sources influencing the subject's gazing behavior, namely given that the subject's gaze is towards X what is the probability that the robot has previously been looking at X[1]? In our interaction this would mean, that the necessity and the sufficiency index for the subject's behavior are computed as follows:

$$necessity_{index} = p(subject'sGazeTowardsX|robot'sGazeHasBeenAtX)$$

$$sufficiency_{index} = p(robot'sGazeHasBeenAtX|subject'sGazeTowardsX)$$

Necessity and sufficiency index are thus non-symmetric. The above description is a computation from the robot's perspective and measures the contingency in the behavior of the subject towards the robot. The overall contingency is then computed as the product of these two variables:

$$contingency = necessity - index * sufficieny - index$$

Note that the value for the contingency lies thus between 0% and 100%, where 100% means perfect contingency (that is a causal relationship) and 0% no contingency at all. The sufficiency in our set up is rising if the tutor is looking at the robot or the object and if the tutor is showing looming behavior. The sufficiency is falling if the tutor is looking somewhere else or is not showing looming behavior.

The necessity in our set up is taking the robots behavior into account and represents the responding behavior of the robot, it is rising if the robot is looking at the tutor or the object and when it points at a object. It is falling if the robot is looking somewhere else or is not showing pointing behavior.
The whole calculation is event driven [5].

III. System Evaluation

In order to investigate the performance of the system we conducted an experiment at University Hertfordshire, in which the participants were asked to present some action to the iCub robot (which we named DeeChee) on this occasion.

A. Participants and Task

Our data set consists of 12 participants who were invited to play with Deechee and asked to come 2 times with a break of one week in between. All participants were native English speakers, with the age range from 21 to 69 years. Most of the participants were students or administrative staff from the University of Hertfordshire. The participants were instructed as follows:

Your task today is to teach something new to the DeeChee. Today and on subsequent days, you will be asked to play with DeeChee. In subsequent sessions with DeeChee, DeeChee may or may not make verbal responses.

- *The DeeChee is equipped with a set of sensors, so that it is connected to our world.*
- *You have a number of coloured boxes with patterns on them in a basket next to you.*

Your job is to play with DeeChee. You are welcome to talk to DeeChee, to use gestures, and you should show the patterns and the colours of the boxes to the robot.
There will be two short tasks for you: I will give you the instruction for the first task now, the task will take 2 minutes. Than I will come back and give you the second task.

- *Your first task : Please present the pattern and the colours of the boxes to DeeChee. In doing this, please make sure to indeed use all the boxes.*
- *The second task : Please teach DeeChee how to stack these different boxes. Please use a different colour and a different pattern for each box.*

B. Setting

The participants were seated across a table looking towards the robot (see Fig.6). The experimenter was sitting behind some monitors to take care of the robot. Three cameras were recording the scene. Participants had the possibility to use three different sized boxes covered with ARToolKit markers.

C. Independent Variable: Contingent behavior of the robot

Based on the model of contingency detection presented in Sec. II a set of manually designed reaction patterns (RP) of the robot were used in the experiment. Using (a) the robot's gazing behavior and (b) the robot's gestures:

[1] note that in the case of mutual gaze the notion "X" is misleading, as in this case the X of the robot would be the subject, whereas the X of the subject would be the robot.

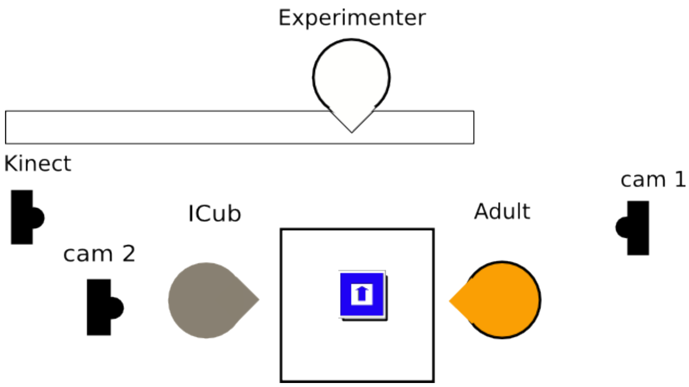

Fig. 6. Setting: The participants were seated on the left side looking towards the iCub-robot, the experimenter was seated behind monitors, three cameras were recording the scene.

- Reaction Pattern 1 (RP-1): system detects ”participant-gazes-at-elsewhere” and reacts by gazing to random locations
- Reaction Pattern 2 (RP-2): system detects ”participant-gazes-at-object” and reacts by directing its gaze at the object
- Reaction Pattern 3 (RP-3): system detects ”participant-gazes-at-robot’s-face” and reacts by directing its gaze to the co-participant
- Reaction Pattern 4 (RP-4): system detects ”participant-points-at-object” and reacts by performing a pointing/looming gesture towards the detected location of the pointing.

When producing the random gaze behavior the robot moves its head-joints randomly but within limits - this stops large/unnatural movements occurring. When the tutor is showing a looming behavior the robot will point at the object for a fixed time limit (about 20sec).

D. Data and Analysis

From the experiment, a set of different data types has been recorded: (a) timeline-based data (video, audio, logging of robot perception and robot internal states) and (b) questionnaires. For the analysis presented here, we combine the different timeline-based data types - audiovisual recordings of the setting and the robot’s logfiles - using the annotation tool ELAN [1]. This enables us to analytically link the system level and the user’s perspective of an interactional event and thereby to close the loop between technical implementation and user studies [12].

In order to evaluate whether the system is able to engage in responsive interaction with the tutor, we use an analytical method that provides insights into the sequential structure of the interaction. Sequential analysis - stemming from Ethnomethodological Conversation Analysis [19], [18] - allows us to investigate the close interrelationship between robot’s and tutor’s actions and how they respond to each other on the level of structural features of the interaction. Important in this approach is the aim to reconstruct the participant’s view (the ”members’ perspective”): We investigate the user’s perception and understanding of the robot’s actions in situ and to which extent they constitute - for the participant - a meaningful, relevant action occurring at an appropriate moment in time, and which further actions they make contingently relevant next.

IV. Results: Participant’s Engagement

In order to evaluate the contingency model and understand its functioning in the concrete interaction with the human tutor, we will compare two cases: (A) one, in which the system is able to engage in a responsive, contingent interaction with the tutor (VP004), and (B) one, in which the system does not do so (VP007). This differentiated view on the system’s performance allows us to study the effects which a contingent vs. a non-contingent robot conduct produces on the tutor’s engagement and presentation of a task. For both cases, we will closely investigate the beginning of the interaction between participant and robot (for the first task) and compare their implications for the tutor’s engagement in the pursuit of the interaction compared (Pitsch et al. 2009 [16]).

A. Contingent behavior (VP004)

This first case (VP004) shows that the implemented contingency model indeed enables the robot to engage in a responsive interaction with the human tutor, in which the tutor’s and robot’s actions respond to each other, on a very fine-grained level of sequential organization, in appropriate and - for the participant - meaningful ways. The first 20 seconds of this interaction allow us to investigate all four implemented reaction patterns (RP) of robot perception and the resulting behavior (cf. section III).

1) Interaction between user and robot: (a) Start of experiment (00.0-12.8 see Fig. 7): Before the actual experiment starts, the iCub robot is placed in its fixed home position facing straight ahead with both arms - bent in a right angle - directed towards the table. When the experimenter starts the system, the robot begins to direct its head to the table, moving head and eyes quickly left and right as if orienting in the environment (#11.9). The participant now orients to the robot, smiles and curiously observes its movements. Then, the experimenter invites the participant to begin the presentation (”you can start”). The robot - which at this stage is choosing its gaze direction randomly (based on its detection of ”participant-gazes-at-elsewhere” - RP-1) - lifts its head and - by incident - turns to the experimenter as if it was orienting to the current speaker (#12.5). Thus, at the moment of first contact, the participant experiences the robot as a reactive system which appears to be able to orient to auditory events in its environment.

(b) Beginning of interaction (12.8-16.2 see Fig. 7): To start the presentation, the participant prepares to grab the cube and briefly gazes towards it (#12.9). With a short delay, the robot detects this conduct as ”gaze-at-object” (13.4, 13.5, 13.7, 13.8). This triggers the robot to also turn its head to the object (#13.6). The system thus performs successfully the behavioral

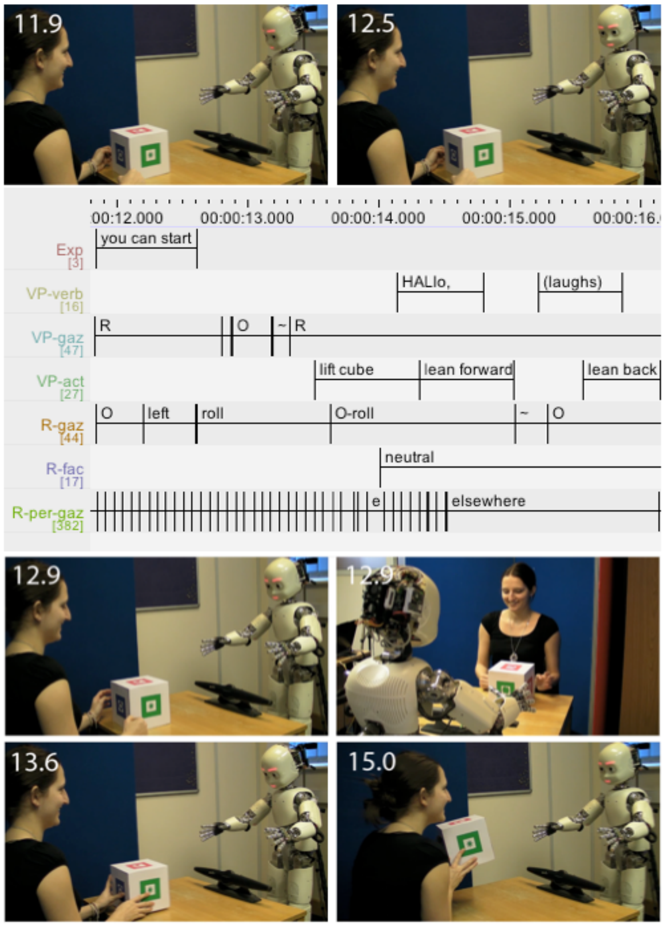

Fig. 7. VP004-Transcript of interaction (seconds 12.0-16.2): Each tier/line represented the annotated conduct of either experimenter (Exp), tutor (VP) or robot (R), and is devided into the following conduct: -verb(=verbal untteranzes), -gaz(=gaze), -act(=actions), -fac(=facelexpresion). The robot's perseption and logfiles are integerated in the tiers R-per-gaz and in some cases R-per-point. each conduct is defined by a start and an end point on the time-line and an annotation value (0 = object, R = robot, ~ = shifting)

pattern provided for in RP-2. – How does the tutor evaluate this robot conduct in and through her own reactions? She lifts the cube (almost up to the level of her face), looks at the robot's face and utters "hello". By greeting the robot, she chooses a social mode of communication for interacting with the system. On a structural level, in human-human-interaction a 'greeting' constitutes the first pair part of an adjacency pair and by virtue of this projects a next action to be relevantly expected from the co-participant (Schegloff [19]). Here, the robot reacts by lifting its head, which leads to a moment of mutual 'gaze' between tutor and robot (#15.0). The tutor acknowledges this with a short laughter both showing receipt of the robot's reaction and a state of amusement and/or astonishment. This way, the participant experiences - before she actually starts the presentation - a robot that appears responsive and attentive to her actions. She begins to treat this machine by using elements of social communicative conduct (greeting), to which the robot responds with a - for her understanding - both appropriate (in terms of human communicational structures) and astonishing (in terms of the situation) reaction.

(c) Explanation - Part 1 (16.2-19.9 see Fig. 8): Next, the participant proceeds with the task of describing the object: She re-directs her gaze to the object and rotates it as to bring it in a position which allows both participants - robot and herself - to look at a particular side (the green cross) and then explains: "so THIS is (-) GREE:N," and points to the cube's green field. At the end of this utterance, she gazes to the robot (#18.3) and thereby addresses this information to the robot. In structural terms, she creates a slot where in human interaction a recipient's acknowledgement is expected. Indeed, the robot reacts by lifting its head, gazing and smiling at the tutor (19.3). This conduct is triggered by the contingency module using RP-3: While rotating the cube, the tutor briefly gazes at the robot's face, which the system detects correctly (16.387, 16.515, 17.215) and launches the gaze-reciprocating behavior. Shortly after this (18.576, 18.860), the system also detects "participant-gazing-at-object", so that its eyeballs start to move quickly between tutor's face and object. – How does the tutor react to this conduct? She waits for about 0.9 seconds, then adds the deictic "HERE," acompagnied by a new pointing gesture to the cube. Thus, she interprets the robot's reaction as appropriate in terms of its timing and the type of action produced. At the same time, she interprets its eye movements as a searching activity to which she provides help for the system to better focus on the relevant location. In this interactional micro-coordination, the tutor treats the system as being responsive on a very fine-grained level orienting to its conduct as sequentially appropriate. Also, she assumes that the system would be able to react on her additional support.

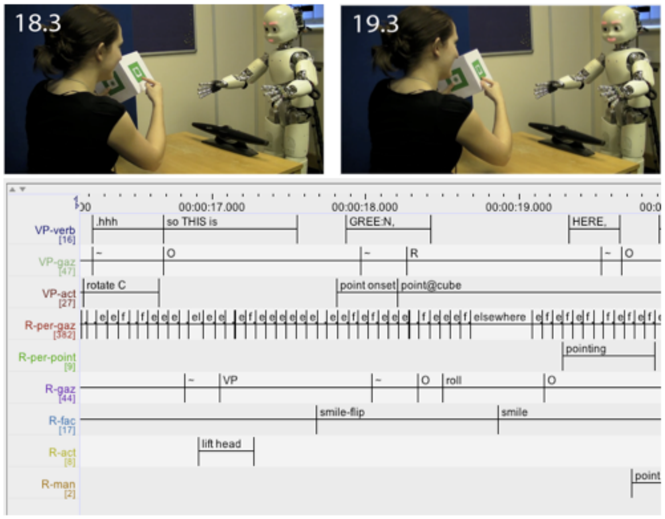

Fig. 8. VP004-Transcript of interaction (seconds 16.0-20.0)

(d) Explanation - Part 2 (19.6-25.0 see Fig. 9): The tutor then continues with her explanation: "and you ca:n, (.) SEE the (-) CROSS in the MIDDLE, " (19.9-22.7) while highlighting with her finger the 'cross' on the cube. At 19.2 the system

detects correctly ”participant-points-at-object”, which triggers the robot’s looming behavior as implemented in RP-4: it points to the object (19.7). In parallel, the system detects ”participant-gazes-at-robot’s-face” (19.1 ff.), which triggers a ’smile’ and gaze towards the tutor, so that tutor and robot achieve a state of apparent co-orientation (22.1). – How does the tutor interpret this conduct? First, she orients to the robot’s emerging gesture by briefly glancing to it (20.2). Second, her utterance is designed as a question (rising intonation), which, again, projects an answer from the recipient. She waits for about 2 seconds, and when the robot produces - now by incident - a head movement that resembles a slow nodding head gesture, she reformulates this as meaning ”YES,” (24.4).

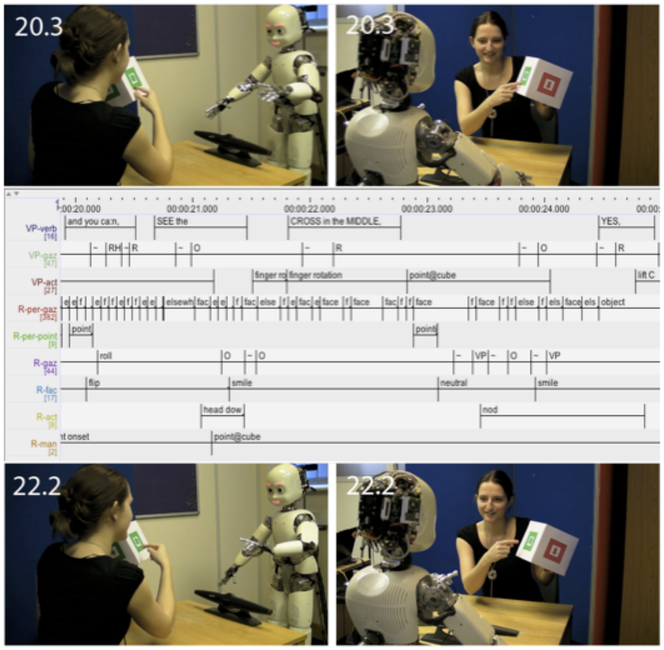

Fig. 9. VP004-Transcript of interaction (seconds 19.5-25.0)

2) Implications for participant’s further engagement: The analysis of the first 25 seconds reveals that the contingency module enables the robot to engage in an interaction with the human tutor, in which not only all four implemented contingency patterns work as assumed, but also - most importantly - the participant orients to the robot’s conduct as appropriate and responsive:

- She explicitly acknowledges the robot’s responsive behavior (laughter)
- She attributes the capability of ”seeing” to the system
- She realizes a form of presentation that is closely oriented towards the robot: She designs her utterances in a way that they project occasions for the robot to produce recipient feedback. Thus, she attributes to the system the ability of being responsive.

This has got implications for the pursuit of the interaction (as shown below in the transcript of her verbal actions): Having experienced the robot as a reactive system, she continuous to present the task in a way as to be highly oriented towards the systems’ actual conduct and display of states and capabilities: She uses short sentences, with a simple repetitive syntactic structure (”and” + S-P-O) and final rising pitch contour, and pauses (ranging between 0.3 and 1.7 seconds) that allow for the robot’s reactions. This way, her presentation is oriented towards the robot, and, at the same time, enables the system to contribute to produce responsive conduct.

```
Transcript VP004 (14.0-40.0):

HALlo,
(0.4)
(laughs)
(0.2)
.hhh so THIS is (0.3) GREEN,
(0.9)
HERE,
(0.3)
and you ca:n (.) SEE the (0.2) CROSS in the MIDDLE,
(1.7)
YES,
(0.6)
on this side it it’s green, (.) ALSO,
(1.7)
and (.) you can see it’s (.) a: SUN,
(1.0)
and eh that’s within a WHITE BO:X,
(0.9)
and (.) then a GREEN BO:X,
(0.8)
a SQUARE,
```

B. Non-contingent behavior (VP007)

In the next section, we will contrast the system’s contingent behavior (VP004) with a case, in which the system is not able to do so (VP007). Here, the examination of the system’s logfiles shows that the system - falsely and persistently - perceives the participant’s gaze as being directed ”elsewhere” (instead of detecting ”participant-gazes-at-robot’s face” or ”-at-object”. In consequence, this triggers the robot’s ’random gaze’ behavior (RP-1).

1) Analysis of interaction between participant and robot: *(a) Start of experiment (00.0-05.7 see Fig. 10):* Differently from the first case, the experimenter starts the system and asks the participant ”you can start now” almost at the same time (00.5 and 03.8), which leads to a situation, in which the robot begins to orient itself in the environment at the same moment at which the participant starts her presentation: When she prepares to grab the object and looks at the robot (05.3), it - by random behavior - begins to turn its head to the left side and withdraws its visual attention (measured as head orientation, 05.6). Thus, the participant’s first impression is a rather non-collaborative robot that appears to disengage once being looked at.

(b) Beginning of interaction and presentation (05.7-10.0 see Fig. 11): The participant then starts the presentation. She utters ”okay”, grabs the object and gazes at the robot, which is still oriented to the side (06.837). While still gazing at the robot (06.7), she pursues with ”so we have got a CU::BE,”. At the

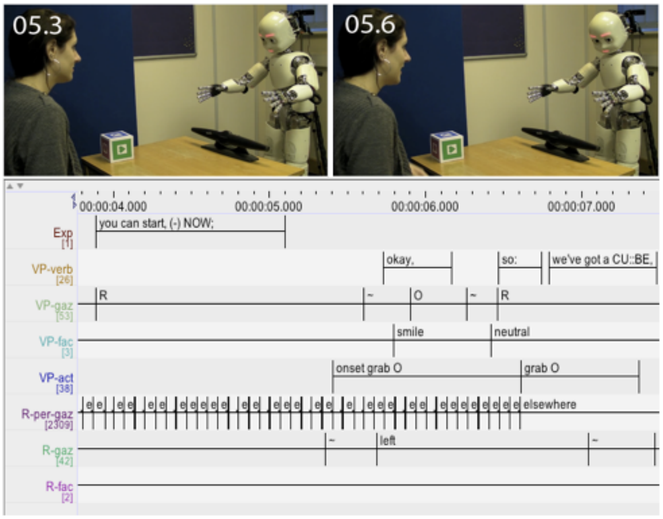

Fig. 10. VP007-Transcript of interaction (seconds 03.8-09.0)

end of her utterance, the robot appears to react by directing its gaze to the ceiling (07.5). Thus, instead of displaying receipt of the information (or at least: a neutral form of no reaction), the robot produces a disprefered action. Yet, the participant continues her presentation with "here in front of us" without receiving any display of the robot's attention.

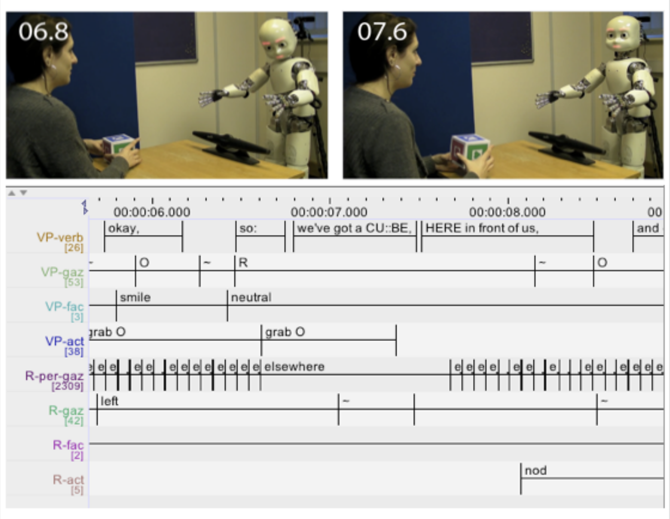

Fig. 11. VP007-Transcript of interaction (seconds 05.5-09.1)

2) Implications for participant's further engagement: The suite of the interaction is characterized by a repetitive reoccurence of this pattern of the robot's withdrawal of gaze at those moments, where the tutor addresses her presentation to the robot and an acknowledging recipient feedback would relevantly be in place. Thus, the robot does not engage in an appropriate responsive interaction with the tutor. This has got implications for the tutor's further engagement. In comparison to VP004, the tutor begins to produce a different conduct towards the system in her tutoring behavior: (i) She visually orients more and more to the object while gazing less at the robot. This means that the robot has lost the opportunity to participate as 'co-participant' in the action presentation (cf. Pitsch et al. 2009). (ii) In her presentation, she uses complex syntactical constructions (S-P-O + relative clause ("which")), with fewer and shorter pauses than the tutor in VP004 (ranging from 0.2 to 0.6 seconds). This not only makes it more difficult for a system to understand the tutor and discriminate actions, but also to limits the possibilities for the robot to give feedback.

```
Transcript VP007 (05.5-34.0):

okay,
(0.2)
so (.) we've got a CUBE here in front of us,
(0.2)
and (.) it's got six SIDES,
(0.5)
and always the same- same dimensions,
(0.3)
and (.) so first we look at the top of the CUBE,
(0.6)
so:: (.) this is a (.) a SQUARE, pasted onto the cube,
which is BLUE,
(0.2)
in the outer SQUARE,
(0.5)
we then have a SMALLER square in the MIDDLE, (.) which
is WHITE
(0.5)
and we have a CONTOUR shape of a MOON,
(0.6)
which shows the blue BACKground (.) through (.) so
that's the TOPside of the CUBE,
```

V. Conclusion

Motivated by recent findings in studies on child-directed interaction, we developed a module allowing for spotting the tutor by monitoring her or his gaze and detecting modifications in object presentation in form of a looming action. We implemented this module onto the robotic platform iCub and conducted a study with participants who were instructed to teach the robot some new objects, their properties and functions. For our study, we hypothesized that the contingency module will improve the interaction with a robot and elicit tutoring behavior from the participants.

For the evaluation of our system, we focused on two participants from the sample and studied their behavior during the very first seconds of their interaction with the robot. For our method we chose the Sequential Analysis as a qualitative approach providing micro-analytical insights into the sequential structure of the interaction and allowing for careful and detailed observations of the interrelationship between robot's and tutor's action and how they respond to each other.

We found that the contingency module enables the robot to engage in an interaction with the human tutor, who orients to the robot's conduct as appropriate and responsive. In contrast, when the robot did not engage in an appropriate responsive interaction, the tutor oriented more to the object while gazing less at the robot. These findings need to be applied to the whole sample of participants in the a second step in order to allow for more general conclusions about the effect and impacts of the contingency module onto the interaction and the users' behavior in general.

Acknowledgment

The work described in this paper was conducted within the EU Integrated Project ITalk (Integration and Transfer of Action and Language in Robots) funded by the European Commission under contract number FP7-214668. All authors gratefully acknowledge the financial support from the FP7 European Project ITALK.

References

[1] H. Brugman and A. Russel. Annotating multimedia/multi-modal resources with elan. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, pages 2065–2068. Citeseer, 2004.

[2] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori, et al. Integration of action and language knowledge: A roadmap for developmental robotics. *Autonomous Mental Development, IEEE Transactions on*, (99):1, 2009.

[3] 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*, 21, 2005.

[4] G. Csibra and G. Gergely. Natural pedagogy. *Trends in Cognitive Sciences*, 13(4):148–153, 2009.

[5] A. Elins. *Object-oriented software development*.

[6] B. Estigarribia and E.V. Clark. Getting and maintaining attention in talk to young children. *Journal of child language*, 34(04):799–814, 2007.

[7] I. Fasel, G.O. Deák, J. Triesch, and J. Movellan. Combining embodied models and empirical research for understanding the development of shared attention. pages 21–27, 2002.

[8] A. Fogel and A. Garvey. Alive communication. *Infant Behavior and Development*, 30(2):251–257, 2007.

[9] L.J. Gogate, L.H. Bolzani, and E.A. Betancourt. Attention to maternal multimodal naming by 6-to 8-month-old infants and learning of word–object relations. *Infancy*, 9(3):259–288, 2006.

[10] H. Kato. Artoolkit. *http://www. hitl. washington. edu/artoolkit/*, 1999.

[11] K. S. Lohan, S. Gieselmann, A. L. Vollmer, K Rohlfing, and B. Wrede. Does embodiment effect tutoring behavior? 2010.

[12] M. Lohse, M. Hanheide, K. Pitsch, K.J. Rohlfing, and G. Sagerer. Improving hri design by applying systemic interaction analysis (sina). *Interaction Studies*, 10(3):298–323, 2009.

[13] S. Machines. faceapi, 2009.

[14] D.J. Matatyaho and L.J. Gogate. Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word–object relations. *Infancy*, 13(2):172–184, 2008.

[15] G. Metta, P. Fitzpatrick, and L. Natale. Yarp: yet another robot platform. *International Journal on Advanced Robotics Systems*, 3(1):43–48, 2006.

[16] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Sussenbach, P. Luff, and C. Heath. the first five seconds: Contingent stepwise entry into an interaction as a means to secure sustained engagement in hri. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 985–991. IEEE, 2009.

[17] K. Pitsch, A.L. Vollmer, J. Fritsch, B. Wrede, K. Rohlfing, and G. Sagerer. On the loop of action modification and the recipient's gaze in adult-child interaction. In *Gesture and Speech in Interaction*, Poznan, Poland, 24/09/2009 2009.

[18] H. Sacks. *Lectures on conversation*. Blackwell, 1995.

[19] E.A. Schegloff. *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge Univ Pr, 2007.

[20] A. Senju and G. Csibra. Gaze following in human infants depends on communicative signals. *Current Biology*, 18(9):668–671, 2008.

[21] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images.

[22] JS Watson. Contingency perception in early social development. *Social perception in infants*, pages 157–176, 1985.

[23] S. Wrede, M. Hanheide, C. Bauckhage, and G. Sagerer. An active memory as a model for information fusion. In *Proc. Int. Conf. on Information Fusion*, volume 1, pages 198–205. Citeseer, 2004.

[24] P. Zukow-Goldring and M.A. Arbib. Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing*, 70(13-15):2181–2193, 2007.