2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. September 9-13, 2012. Paris, France.

[PRE-PRINT]

The final publication is available in the Proceedings of Ro-Man 2012, p. 974-981
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6343876&tag=1

Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human-robot-interaction

Karola Pitsch, Katrin S. Lohan, Katharina Rohlfing, Joe Saunders, Chrystopher L. Nehaniv, Britta Wrede

Abstract— **The paper investigates the effects of a robot's "on-line" feedback during a tutoring situation with a human tutor. Analysis is based on a study conducted with an iCub robot that autonomously generates its feedback (gaze, pointing gesture) based on the system's perception of the tutor's actions using the idea of reciprocity of actions. Sequential micro-analysis of two opposite cases reveals how the robot's behavior (responsive vs. non-responsive) pro-actively shapes the tutor's conduct and thus co-produces the way in which it is being tutored. A dialogic and a monologic tutoring style are distinguished. The first 20 seconds of an encounter are found to shape the user's perception and expectations of the system's competences and lead to a relatively stable tutoring style even if the robot's reactivity and appropriateness of feedback changes.**

I. Introduction

Following the "social learning" paradigm, the interaction between human and robot gains importance for robotic learning [1]: A human tutor presents some action to the robot, which is supposed to observe it, understand its structure and reproduce it. The quality of the tutor's presentation is thus crucial. However, the precise ways in which tutors present some action differ considerably on the micro-level, and are thus challenging to discriminate. While one strand of research develops methods for detecting and processing the tutor's conduct, we suggest to explore the ways in which the robot could exploit the interactional situation: Inspired by tutoring in adult-child-interaction [2], we argue that a robot's "on-line"-feedback can influence the tutor's presentation in the moment it is emerging. The robot could pro-actively help generate the input from which it would benefit most.

In this process, the beginning of an encounter plays a central role. During the first moments, the participants get in contact with each other, identify and categorize the co-participant and build expectations about the ensuing interaction [3, 4]. When a tutor presents some action to the robot for the first time, he/she does not know what the system is able to do, what it understands, and what it might be reactive to. Thus, while beginning to present an action, participants also have to explore the system's competences. The user's experience of a robot as responsive vs. non-responsive in the firsts moments of an interaction has implications for its pursuit [4].

In this paper, we investigate how the robot conduct during the first seconds of a tutoring encounter shapes the tutor's expectations and the ways in which the presentation is carried out. We present findings from an HRI-study, in which a robot was equipped with a module that autonomously generates feedback (gaze, pointing) based on the system's perception of the tutor's actions [5]. We examine:

1. How does the robot's conduct influence the ways in which the tutor presents an action? How do tutor and learner jointly establish the tutor's presentation format through their micro-coordination in the first moments of the encounter?

2. What are the implications of the first impression for the pursuit of the interaction? If the robot is experienced as responsive at the beginning, how does the tutor treat occurring non-responsive conduct at a later stage? How does he/she treat responsive conduct of a robot, which has been experienced as non-responsive at the beginning?

This way, we also suggest an interactional and emergent perspective on user expectations [6]. We shed some light on the question how they are produced and updated step by step.

II. Feedback in Tutoring Situations

Feedback for technical systems is important in HCI/HRI to enable the user to understand the system's internal states and whether some input has been received. However, studies on robotic "social learning" have only scarcely investigated this dimension. Tutoring in HRI has been conceived of as a one-way communication, in which the robot observes the tutor's actions without actively taking part in the social situation. In the few cases, in which a dialogic perspective is taken and the robot provides feedback, it has been designed to produce a positive/negative statement *after* the tutor has finished his/her presentation [e.g. 7]. Tutors acknowledged the robot's feedback, but it needed to be more informative.

Recent HHI-studies on parent-infant tutoring suggest that – based on constant monitoring – the tutor adjusts his presen-

*This research was conducted within the EU-project 'iTALK', FP7-ICT-214668) of which all authors acknowledge the financial support. Additionally, Karola Pitsch acknowledges the financial support from the Volkswagen foundation (Dilthey fellowship).

Karola Pitsch and Britta Wrede are with the Applied Informatics Group and the Research Institute for Cognition and Robotics (CoR-Lab), Faculty of Technology, Bielefeld University, 33501 Bielefeld, Germany (correspond-ing author: karola.pitsch@ uni-bielefeld.de).

Katrin S. Lohan is with the Cognitive Humanoids Lab, Istituto Italiano di Tecnologia for Robotics Brain & Cognitive Sciences, 16163 Genova, Italy.

Katharina Rohlfing is with the Emergentist Semantics group at Bielefeld University, CITEC, 33501 Bielefeld, Germany.

Joe Saunders and Chrystopher L. Nehaniv are with the Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield AL10 9AB, United Kingdom.

tation on a micro-level to the learner's needs [2, 8]. The infant learner's actions provide feedback about her state of attention and/or cognitive involvement both (i) "turn-by-turn" (i.e. after an utterance/action) and/or "on-line" (i.e. during the tutor's emerging presentation) [9]. With regard to "on-line" feedback an "interactional loop" between the tutors' hand motions and the learner's gaze has been shown for a stacking cups scenario: When presenting the action, the tutor attempts to guide the infant's visual attention by adjusting the movement of her hand. Vice versa, the learner's gaze (following/anticipating the action, disorienting) pro-actively shapes the emerging trajectory of the tutor's hand [2].

Based on these observations, we suggest that the robot's "turn-based" *and* "on-line" feedback are important for the tutoring situation: The robot has at its disposal a resource with which it might be able to shape the tutor's conduct as to best benefit from the presentation. What exactly participants react to, how they interpret the robot's conduct, and which strategies might be best is a matter of empirical research.

III. ROBOT SYSTEM

We equipped an iCub robot with a feedback module [5], which is based on reciprocity of actions and realizes an observation-production-loop. The system observes (a) the tutor's gaze classifying it into three categories (to learner, to object, elsewhere) and detects (b) pointing gestures. It uses as sensors the robot's internal cameras, an external Kinect, a microphone and object tracking (ARToolKit). No learning algorithms are integrated. The robot's reactions respect a 200-300 ms time-frame [5]. To robot's reciprocation is based on the following four manually designed reaction patterns:

- Reaction Pattern 1 (RP-1): system detects "participant-gazes-at-elsewhere" and reacts by gazing to random locations; this is accompagnied by a neutral robot face.
- Reaction Pattern 2 (RP-2): system detects "participant-gazes-at-object" and reacts by directing its gaze at the object; this is accompagnied by a smile on the robot's face.
- Reaction Pattern 3 (RP-3): system detects "participant-gazes-at-robot's-face" and reacts by directing its gaze to the co-participant; this is accompagnied by a smile on the robot's face.
- 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.

IV. STUDY AND DATA

We conducted a user study, in which 11 participants (native English speakers, university students and untrained users) were given two tutoring tasks to be explained to the iCub (named DeeChee). Here, we focus on the first task (3 minutes duration), for which they were provided with three differently sized boxes with colored patterns. They were instructed as follows: "Please present the patterns and the colors of the boxes to DeeChee. In doing this, please make sure to use all boxes." The participants were informed that the robot was equipped with sensors and that they could talk and use gestures when carrying out the task.

Participant and robot were seated face-to-face across a table. The experimenter was present in the room at all times, sitting behind a computer screen mostly invisible (Fig. 1). After instructing the participant, the experimenter started the system. The interactions were videotaped with two HD-cameras, the robot's perception and internal states were logged and synchronized afterwards using ELAN.

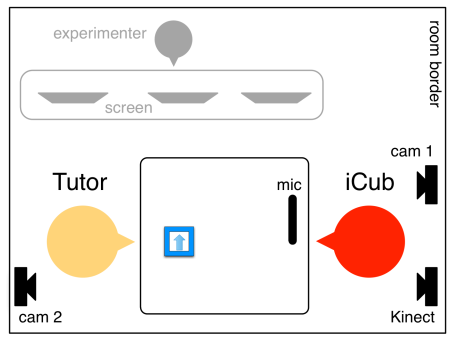


Figure 1. Set up of the study

As the system runs autonomously, it is faced with the insecurities of robot perception. This results in cases, in which the system generates responsive conduct and those in which it does not. This allows us to investigate the effects, which a responsive vs. a non-responsive system has in tutoring situations. Here, we focus on two opposite cases.

An evaluation – linking the system's and the user's perspective – revealed that the perception-action-loops work and that the robot's non-/responsive feedback *co-occurs* with different tutoring styles [5]. Here, we investigate (1) *how* the tutor's presentation is shaped through the robot's feedback; and (2) the *implications of the user's first impression*.

V. METHOD

We use an analytical method that provides insights into the sequential structure of the interaction: Ethnomethodological Conversation Analysis (EM/CA) [10]. This allows us to investigate the interrelationship between robot's and tutor's actions and how they respond to each other on the structural level. Important is the aim to reconstruct the participant's view ("member's perspective"): We investigate the user's perception and understanding of the robot's actions and to which extent they constitute – for the participant – a meaningful, relevant action occurring at an appropriate moment.

EM/CA's way of proceeding is qualitative. Case analyses are undertaken to reveal – from the data themselves – the analytical issues and categories. It consists of manual analysis, i.e. repeated inspection of video-data and transcribing the interaction (see appendix) to uncover the timing and relationship of the actions. Its goal is to find the structural organization and how one action makes another one contingently relevant next. On this structural level also the absence of an (otherwise expected) action can be accounted for.

VI. Responsive Beginning Leads to Dialogic Tutoring Style and Normalization of Robot's Conduct

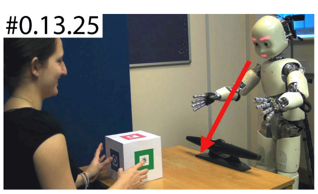
We start the analysis with a first case (participant 04), in which the feedback module works as intended and reciprocates the tutor's conduct. This leads to a start of the encounter, in which the robot appears responsive to the tutor's actions. How is this achieved? How does this lead to a particular form of tutoring? What are further implications?

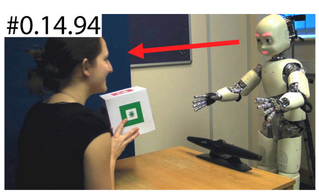
A. Responsive Beginning and First Task

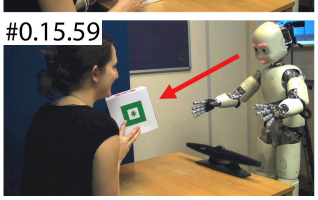
1) First contact: Experiencing an active system

When the experimenter (E) starts the robot (R), the iCub firstly moves its head and eyes randomly, then looks down to the table. The tutor (T) experiences thus an active system, which orients to a relevant location (#00.13.25). The tutor then lifts the box, gazes and smiles at the robot and addresses it with "hello" (02). The iCub – detecting the tutor's 'gaze-at-robot' and accordingly producing RP-2 (for details on the robot's perception see the examples in [6]) – reacts by lifting and orienting its head (which appears as 'gaze') to the tutor's face (#00:14.94). She leans forward and thereby also brings the box further towards the robot and briefly gazes at it. Again, the system recognizes this short gaze shift and reciprocates it by directing its head to the object (#00.15.59).

```
Fragment 1 (00:00–21:00): Beginning & box 1/1st
side

01 E:      |okay, |(7.0)   |you can start,
   T-gaz:  |@E     |@R  |@E |@R
   R-act:  |moves head & eyes …

02 T:      |(1.5)    |hello, |(0.5) (laughs)
   T-gaz:  ...
   T-act:  |grabs O |lift O
   R-gaz:  |@Table            |@T       |@O

           |#0.13.25          |#0.14.94 |#0.15.59
```







This way, a sequence of well-timed actions ensues, in which the robot appears to be reactive to the tutor's conduct and to be relevantly orienting between the tutor and the object. The tutor 'comments' this with a short laughter (02).

2) Explanation of 1st side – I: Integrating linear robot conduct into interactional sequence structures

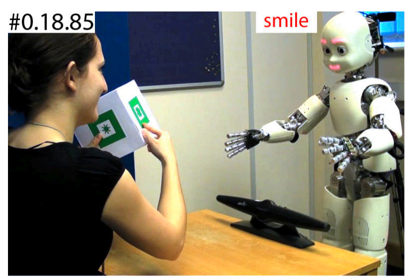
The tutor then starts to explain the box and its patterns to the robot: "so THIS i:s GREE:N," (03). During this utterance, the robot seems to closely follow her presentation:

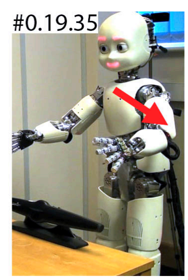
(i) The tutor initially looks at the box (03: @O), and the robot reciprocates her gaze to the object (03: @O) (RP-2).

(ii) Following the verbal deictic "THIS", the robot lifts briefly its head (which is rather an unmotivated 'noise' in its behavior) (03: head↑), then reorients to gaze at the object.

(iii) While uttering her last word "GREE:N," the tutor points to the box and marks her utterance as completed (rising intonation, syntactical structure) which she addresses via gaze to the robot. Then she briefly pauses and thus creates a structural slot where some next action of the co-participant would be expected (display of understanding, acknowledgement etc.). The robot detects the tutor gazing to the robot (@R) and reacts by producing a smile at the end of "GREEN," (#0.18.85). The smile appears just in the sequential position, at which the co-participant's reaction would be relevant and thus could be interpreted as a form of response.

```
03 T:      |so THIS |i:s   |GREE:N, (0.3) |
   T-gaz:  |@O              |@R
   T-act:                   |point
   R-gaz:  |@O≈      |head↑|@O≈
   R-fac:                   |smile-flicker |smile
                                           |#0.18.85
```





This way, the robot appears to be closely monitoring and following the tutor's presentation and eventually producing a "response" at a structurally relevant position. The "smile" as a reaction is generated as part of RP-2, i.e. as a reaction to the tutor's gaze at the object. However, the tutor integrates the "smile" into the sequential structure at the place where a response is projected. Thus, the robot's linear conduct co-occurs – accidentally – with specific sequential structures.

Then, the robot's gaze shifts – rather unmotivated – to its left side (#0.19.35). The tutor reacts by adding "HERE" and attempting to repair the robot's visual focus of attention.

```
04 T:      |(0.4) |HERE,
   R-gaz: |≈≈≈|@left
               |#0.19.35
```

This shows that the tutor – after the initial adjustments of the robot's conduct to her actions – orients to fine details of the robot's shifting orientation. The tutor appears to perceive the robot as responsive and producing interactional moves. She assumes that it could react to her additional support.

3) Explanation of 1st side – II: Establishing an interactional format for presenting. The dialogical style

A similar structure for the tutor's next utterance occurs: "and you ca:n (.) see the cross in the middle,".

(i) The tutor gazes and points to the object (05: @O). The robot reciprocates by gazing and pointing to it (#0.20.68), which is noticed by the tutor's glance to the gesture.

(ii) At the end of her utterance, the tutor's intonation raises, she looks to the robot and pauses for a second. Her utterance is thus designed as a question and, again, opens a slot, which projects a co-participant's answer. The robot reciprocates the gaze, so that they achieve a state of apparent co-orientation. At this moment, the robot produces an upward-downward head movement, which accidentally *looks* like a head nod (but is not programmed). The tutor waits for another second, then reacts with "yes" (03) – thereby treating the head movement as a nodding gestures and as an affirmative response and display of understanding.

```
05 T:      |and |you |ca:n (.) |see the cross |in
   T-gaz: |@O        |@R        |@O             |@R
   T-act:  ...
   R-act:       |point
   R-fac:  ...                  |
                                |#0.20.68

06 T:      the middle, (1.0) |                |yes,
   T-gaz: @R ...
   R-act:                     |point + head↑↓ |
```

Note that the robot was not programmed to produce any head nods. It is rather the tutor who interprets the robot's conduct as a nod and treats it as a locally meaningful action. Both the "smile" in line 03 and the apparent "nod" in line 05 are forms of conduct, which the robot produces at each moment for the first time. And it produces them in a particular structural position with regard to the tutor's tutoring: at the moment at which a robot reaction is expected. This way, an interactional format begins to emerge, in which the tutor addresses the robot at the end of her utterance, pauses for a short instance, and the robot produces some sort of reaction (such as "smile" or "nod"). The tutor thus interprets the linearily generated "smile" and the accidentally produced "nod" as actions filling the projected slot of 'acknowledgement' of a presentation. Over the presentation of the next two sides of the box, a similar structure can be found, in which the tutor interprets the robot's head movements again as "nodding" and thus as display of understanding at a structurally relevant point.

This way, an interactional format for presenting the actions emerges: a **dialogical style** (Fig. 2). It has got the following features: The tutor produces short sentences with a simple repetitive syntactical structure, a final-rising pitch contour and gaze directed to the robot at the end of the utterance. The robot appears to be closely monitoring and following the tutor's gaze orientation and provides some sort of acknowledgement at the end of the tutor's utterance.

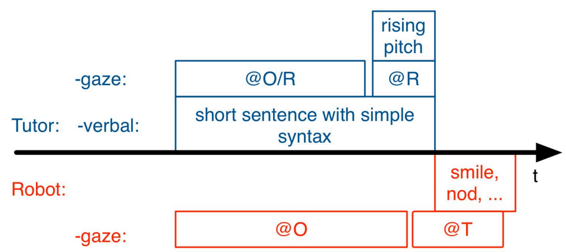


Figure 2. Interactional presentation format: Tutor's dialogical style

The tutor's presentation style becomes conceivable as a co-construction between both tutor and robot: The tutor produces a small utterance-package, gazes to the robot and waits for its reaction (i.e. opens a structural slot for the robot to show a reaction resp. display its understanding). The robot's conduct, which occurs at that moment, is interpreted by the tutor as a relevant action and treated as meaningful. This way, human and robot begin to collaboratively produce the style and the interactional format of the tutor's presentation. Both participants provide a set of actions, which react upon the other's conduct. However, the tutor undertakes an additional – and crucial – sense-making effort. She integrates the robot's linear conduct (i.e. a simple response-pattern) into a sequential action structure.

B. Continuation of the interaction: Robot's non-coherent conduct is 'normalized' by the tutor

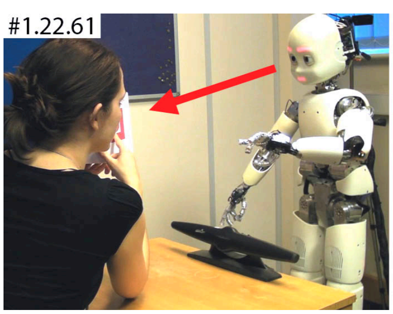
During the presentation of the next sides of the boxes the robot's conduct remains systematic and responsive. Does the tutor's presentation style remain consistent over the teaching period? Also if the robot becomes unresponsive? Does the tutor's interpretation of the robot's conduct change?

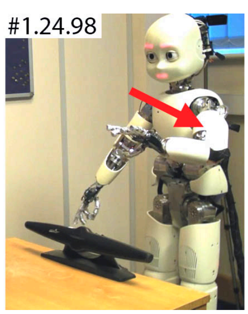
In the pursuit of the interaction, a couple of moments occur, at which the robot's conduct does not appear reactive to the tutor's actions: During the presentation of the 4th side, it gazes away twice when being addressed at the end of the utterance. It is during the presentation of the 5th pattern, the tutor begins to visibly react to this conduct. She explains: "and on this side, it's RED," (01) gazing to the robot at the end of the utterance (@R); she continuous with "and we have got a red box, (.) and then a white box" (02-03, @R) also gazing to the robot. During this, the robot's gaze alternates between being oriented firstly to the table, then to the tutor and finally to the object (@table, @T, @O) – thus, looking at relevant positions, but not appearing to 'respond' at the end of the tutor's units.

```
Fragment 2 (01:12 – 01:28): Box 1, 5th side

01 T:         |and on THIS side, it's |RED,
   T-gaz:  @R |@O                     |@R
   R-gaz:  @table                |@T  |@O
   R-act:  point@O

02 T:      and we have a |red |box, (.)
   T-gaz:  @O            |@R  |@O
   R-gaz:  ...  @table    |@T  |@O

03 T:      and       |then a |white |box, (0.5)
   T-gaz:                    |≈≈≈≈≈ |@R
   R-gaz:  @table    |↑↑      |≈≈≈|@O
   R-fac:  smile
```

When the tutor continuous: "with a moon inside" (04) again gazing to the robot at the end of the utterance (@R, #01.22.61). However, the robot directs its head to its left side, (≈≈left, #01.24.98): it appears to gaze away.

```
04 T:      with a |moo:n |inside,   |
   T-gaz:  ...    |@O    |@R
   R-gaz:  ...                      |≈≈left
                         |#01.22.61|#01.24.98
```





This time, the tutor reacts by reformulating her last utterance: "the shape of a moon" (05). As such reformulations have not been found in her presentation so far, it is a remarkable conduct and can be seen as a reaction to the robot's preceding disorientation.

```
05 T:      the |shape of |a moo|:n,
   T-gaz:  @R            |@O   |@R
   R-gaz:      |@table ...

06 P:      and that's red,
   R-gaz:  ...
```

This example shows that the tutor indeed reacts to the robot's changing conduct on a fine-grained level of verbal formulation strategies. However, the dialogic presentational style remains stable over the pursuit of the interaction (further empirical evidences cannot presented here due to space constraints). Thus, the tutor normalizes the robot's unsystematic conduct and integrates it into her presentation.

It seems that tutor and robot have build up a routine of presentation at the beginning of their interaction, which is pursued later on. Further empirical examples show that, later on, only occasional moments of responsive robot conduct suffice for the tutor to be treated as 'anchors' that she could integrate as meaningful actions into the sequential structure of her presentation. On the one hand, her presentation style – which provides moments to closely observe the robot's conduct and to provide time for its (seemingly meaningful) reactions – enables her to directly react to the robot's conduct. On the other hand, the tutor has build expectations about the robot's competences during the responsive beginning of the encounter, which now come into play. Thus, we find that a tutor – who has experienced the robot as responsive during the beginning of the encounter and has build up certain interactional routines with the robot as co-participant – appears 'permissive' once the system starts to also produce unexpected non-coherent conduct. Occurring 'failures' in responsiveness appear to be normalized as long as the tutor subsequently also finds elements in the robot's conduct that it could interpret as meaningful and integrate into the sequential structure of her presentation.

VII. Non-Responsive Beginning of Interaction Leads to Persistent Monologic Tutoring Style

We will now explore a contrasting case (participant 07), in which the system does not reciprocate the tutor's actions and appears to act rather randomly. How does this lead to a particular form of tutoring?

A. Non-Responsive Beginning and First Task

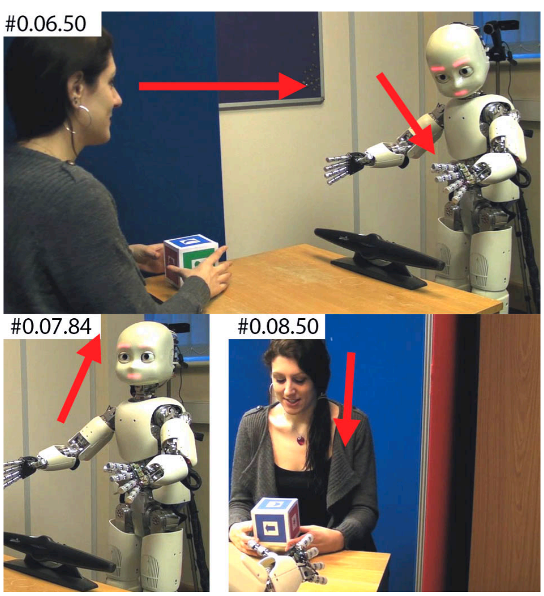
1) First contact and introduction of the box: Experiencing a robot orienting to random locations

Participant 07 has got comparable starting conditions to participant 04: While the tutor is already sitting at the table facing the iCub, the experimenter starts the system. Before this, she can observe the robot moving its head randomly for about four seconds and then looking at the table. Thus, the robot appears active, orienting to a relevant location.

When she starts to present the first box ("okay (.) so w've got a CUBE here in front of US,"), she closely observes the robot (01-02, @R). However, the system does not detect "gaze-at-robot" and thus does not reciprocate the tutor's gaze (as suggested with RP-3). Instead the iCub continuous to gaze sideways-down to the table (#0.06.50), then lifts its head to look towards the ceiling (02: @↑↑, #0.07.84). Neither are these gaze directions relevant for the tutoring situation, nor do they occur at a sequentially relevant moment of the tutor's presentation. The tutor thus experiences a non-responsive and not task-related robot. At the end of her utterance – which is designed as a short syntactical unit with rising intonation similar to participant 04 – she redirects her gaze to look at the object and continuous with a hesitation (03: "euhm", #0.08.50). This way, she does not – contrary to participant 04 – open a slot where some robot reaction would be relevantly expected.

Fragment 3 (00:00-20:00): Beginning and box 1/1st pattern

```
01 T:      okay, |(.)
   T-gaz:  @O    |@R
   R-gaz:  @table-left

02 T:      so w've got a CUBE here in front of|US,
   T-gaz:  ...                                |@O
   R-gaz:  ...             |≈≈≈≈≈ |@↑↑
           |#0.06.50          |#0.07.84     0.08.50
```



She continuous with a next utterance: "it's got six SIDES," (03). During this, she re-orients to the robot (@R) and sees that the iCub now indeed relevantly gazes to the object (RP-2 after having picked up the tutor's "gaze-to-object" from line 03). But directly after she has begun to look at the robot, the iCub starts to somewhat bizarrely rotate its eyes (03: "rotate"). She reacts by redirecting her gaze to look at the object while the robot then orients to the table. This way, no seemingly mutual gaze or joint attention is established between tutor and robot. Also, again at the end of the utterance, the tutor does not project any reaction from the robot: she gazes to the objet (03: @O) and fills a potential pause with the hesitation "euhm" (04).

```
03 T:      euhm,|(.) |it's |got |six |SIDES,
   T-gaz:       |≈≈≈ |@R        |@O
   R-gaz:  ≈≈≈≈≈|@O        |rotate   |@table-left
```

A similar interactional pattern occurs for the next utterance: "all of the same dimensions" (04). Again, the tutor gazes to the robot at the beginning (@R), which – in turn – redirects its gaze from looking at the left side of the table to the right side. Again, this incites the tutor to redirect her gaze to the object (@O).

```
04 T:      euhm all |of |the same |same di|mension
   T-gaz:  ...      |≈≈≈|@R       |@O
   R-gaz:               |≈≈≈|@table-right |rotate
```

2) Explanation of 1st side – I: Robot gazes away at end of units which leads to extension of syntactical units

When she then starts to explain the box's first side (05-07) she uses a presentation format similar to the one suggested by participant 04: short utterances which are directed to the robot via gaze (05 and 07: @R) leaving a short pause (05, 07), in which a potential robot reaction would be relevant:

```
05 T:      euhm so |FIRST, (.)
   T-gaz:  ...     |@R
   R-gaz:  @left

06 T:      |we look at the |TOP of the cube,
   T-gaz:  |@O
   R-gaz:  ...             |@O

07 T:      so::, this is a SQUA|RE, (.)
   T-gaz:  ...                 |@R
   R-gaz:  ...
```

However, the robot only gazes to its left side resp. the object, is neither attentive nor reactive to her actions and does not fill the slot where a reaction would be expected.

During the next utterance, the tutor adjusts her syntactical structure to the lack of the robot's responsiveness. She addresses the unit "pasted onto the cube, which is blue," to the robot via gaze (08-09). Seeing that the robot continuous to gaze to its left side (@left) instead of providing a reaction, she continuous by adding "in the outer square," (10).

```
08 T:      |pas|ted |onto the cube,
   T-gaz:   ...     |@O
   R-gaz:  |≈≈≈|@left
```

```
09 T:      which is |blue,
   T-gaz:           |@R
   R-gaz:  ...

10 T:      in the |outer square, (.)
   T-gaz:  ...    |@O
```

Her originally short, syntactically complete units (similar to those of participant 04) are thus extended as a result of the robot's lack of attention resp. response. Also here, the robot co-constructs the format of the tutor's action presentation.

3) Explanation of 1st side – II: Establishing an interactional format for presenting. The monologic style

Then, a new pattern in the tutor's presentational style emerges. Instead of looking at the robot at the end of an utterance and thereby projecting a slot for a response or reaction, she begins to organize her gaze differently: She looks at the object during the presentation (@O), briefly looks up to the robot *before* the end of the unit (@R) and reorients to the object at the end of her utterance (@O).

```
11 T:      we then have a |smaller square |in the
   T-gaz:  @O
   R-gaz:  ...            |rotate         |@left

12 T:      midd|le,|
   T-gaz:  ... |≈≈ |
   R-gaz:  ...

13 T:      which |is WHITE,
   T-gaz:  @R    |@O
   R-gaz:  ...
```

Her gaze behavior allows to briefly check the robot's attention. But it does not allow to constantly monitor its actions nor does it provide a slot for an active contribution or response from the co-participant. She displays that she does not expect a reaction from the robot any more at sequentially relevant moments. Thus, the robot's non-responsive feedback and the lack of micro-organization between tutor and robot have step-by-step shaped the tutor's presentational style. From a dialogic beginning similar to the one used by participant 04, the tutor proceeds to a **monologic style** (Fig. 3) with the following features: Extended complex syntactical units, a gaze behavior which only briefly checks on robot's attention during the utterance, but does not allow to monitor its actions neither provides a slot where some robot reaction is expected.

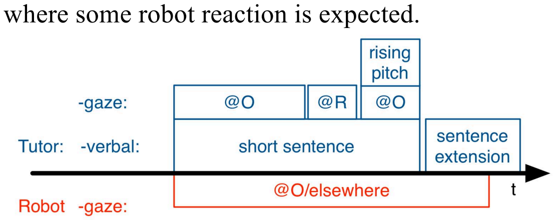


Figure 3. Interactional presentation format: Monologic tutoring style

B. Continuation: Seemingly responsive actions are not oriented to by the tutor

Given the tutor's emerged monologic presentation style, the robot does not have – in the pursuit of the encounter – any systematic structural opportunities at which it could produce relevant responsive feedback to the tutor's actions. Thus, even if the robot, at times, generates some conduct which (from an external observer's perspective) appears to be responsive to the tutor's actions, it is not oriented to and interpreted as meaningful by the tutor.

In the following fragment, the robot produces a downward-upward head movement (05) after the tutor's completed syntactical unit: "it's a six sides three-dimensional shape," (04-05). The robot's action is comparable to the one exhibited during the interaction with participant 04, which was interpreted by the tutor as the meaningful action of "nodding" (see section VI.A.3).

```
Fragment 4 (01:50 – 02:06):

01 T:      HERE, (.) we've got a LARGER cube,
   T-gaz:  @O

02 T:      so:: (.) the pattern is the SAME,
   T-gaz:  ...

03 T:      it's a six sides three-dimensional
   T-gaz:  ...

04 T:      shape,
   T-gaz:  ...

05 T:      euhm (.) and this time on the top,
   T-gaz:  ...
   R-act:  ↓↑

06 T:      we've got a RED square,
   T-gaz:  ...
```

However, as the tutor has stopped looking at the robot and waiting for its reactions, she does not realize this action. This shows, on the one hand, that the gaze pattern of the emerged monologic presentation style does not allow the tutor to realize eventually later occurring responsive robot conduct. On the other hand, we see that the tutor has build a perception of the robot and expectations about its conduct which categorize the system as non-responsive and not taking part as co-participant in the tutoring interaction. Thus, once a robot system has been conceived of as non-responsive, 'normal' responsive conduct would not be sufficient for the tutor to change his conduct. Rather the system would need to exhibit more powerful means. Thus, the robot would better get its reactions right at the beginning of the encounter if we wanted it to be treated by the user as a co-participant in an interactive exchange.

VIII. Conclusion and Implications

Starting from the observations in HHI according to which the learner can actively shape the tutor's presentation through his/her "on-line" feedback, we investigated this relationship for a tutoring scenario in HRI. Presenting a qualitative micro-analytic investigation of two opposite cases, we were able to show that also a robot can influence through its gaze behavior, head movement and pointing gesture the concrete ways in which the tutor presents some action:

- A dialogic tutoring style (Fig. 2: short sentences with simple (S-V-O), repetitive syntax, final-rising pitch contour, gaze directed to the robot at the end of the utterance, pause after a syntactical unit) can be distinguished from a monologic tutoring style (Fig. 3: extended complex syntactical units, gaze only briefly checks on robot's attention during the utterance, no structural slot provided for robot reaction).
- The robot's responsive vs. non-responsive conduct co-constructs this specific style. While in both cases, the tutor started out with a robot-oriented way of presenting, the emergence of these two styles during the first 20 seconds of the encounter could be traced.
- When the robot appears responsive at the beginning, it has good chances of being presented with a dialogic tutoring style, which also can remain active if the system occasionally fails later on. When the robot does not appear responsive at the beginning, it is likely to be confronted with a monologic presentation, even if the system becomes more responsive later on. Thus, to provide robot systems with means of engaging in (crucially: initial) responsive conduct is important to bring the robot in a position to be tutored in a robot-oriented dialogic manner.

These results lead to implications on a more general level:

- Firstly, they pave the way towards an interactional view on robotic "social learning". The robotic learner should not be conceived of as a passive observer of some action presentation. Rather, it is immersed in the social situation. It can – through its own conduct – pro-actively shape *how* the tutor presents the action. Here, the dialogic tutoring style provides the benefit of well-packaged units and a dedicated slot for learner feedback. Thus, if we were to design concise feedback strategies (including gaze, head and hand gestures to signal state of attention and understanding), the robot could participate in generating the input from which it would benefit most.
- Secondly, the micro-analytic perspective shows how user expectations emerge in the moment of the interaction. They are shaped during the first few seconds of an encounter, and seem to remain relatively stable during the pursuit of the interaction.

The current results are based on two clear opposite cases taken from a corpus of 11 participants. The other interactions show cases 'in between' and will be detailed in relation to the cases presented here in a follow up paper. The results are limited in that they explore only three minutes of HRI. The shaping of user expectations and their stability would need to be traced over a longer period of time in further studies [11]. Subsequent investigations on the effects of robot "on-line" feedback should follow detailing the effect of different resources and their interest for machine learning approaches.

Appendix: Transcription Conventions

Each line gives the conduct of experimenter (E), tutor (T) or robot (R), their verbal utterances, actions (-act), gaze (-gaz) or facial expressions (-fac). The GAT convention are use for verbal utterances (in general lower case spelling; upper case for stressed syllables, punctuation gives prosodic features (',' = rising; ';' = falling). Important annotation symbols are: O = object, R = robot, T = tutor; @ = at; ≈ = shifting. Video stills are linked to the transcript via time code.

Acknowledgment

The authors thank Hagen Lehmann and Christian Dondrup for their support in conducting the study and Lukas Rix and Raphaela Gehle for transcribing/annotating the data.

References

[1] B. Wrede, K. Rohlfing, M. Hanheide and G. Sagerer (2009). Towards learning by interacting. In: HRI 2008, 139-150.

[2] K. Pitsch, A.-L. Vollmer, J. Fritsch, B. Wrede, K. Rohlfing and G. Sagerer (2009). On the loop of action modification and the recipient's gaze in adult-child-interaction. In: GESPIN 2009, 1-7.

[3] E.A. Schegloff (2002). Opening Sequencing. In: J. E. Katz and M. Aakhus (Eds.), Perpetual Contact: Mobile communication, private talk, public performance, Cambridge University Press, 326-385.

[4] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Süssenbach, P. Luff and Ch. Heath (2009): "The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in Human-Robot-Interaction. In: RO-MAN 2009, Toyama, Japan, 985-991.

[5] K.S. Lohan, K. Rohlfing, K. Pitsch, J. Saunders, H. Lehmann, C. Nehaniv, K. Fischer and B. Wrede (2012). Tutor spotter: Proposing a feature set and evaluating it in a robotic system. In: International Journal of Social Robotics, 4(2), 131-146.

[6] K. Pitsch and B. Koch (2010): How infants perceive the toy robot Pleo. An exploratory case study on infant-robot-interaction. In: Proceedings of the Second International Symposium on New Frontiers in Human-Robot-Interaction (AISB), Leicester, UK, 31/03/2010, 80-87.

[7] A. Alissandrakis and Y. Miyake (2009). Human to Robot Demonstrations of Routine Home Tasks: Acknowledgement and Response to the Robot's Feedback. In: AISB 2009, 9-15.

[8] P. Zukow-Goldring (2012). Assisted imitation: First steps in the seed model of language development. In: Language Sciences.

[9] A.-L. Vollmer, K. Pitsch, K.S. Lohan, J. Fritsch, K. Rohlfing and B. Wrede (2010) Developing feedback: How children of different age contribute to an interaction with adults. In: ICDL 2009, 1-9.

[10] Schegloff E. (2007) Sequence organization in interaction: A primer in conversation analysis. Cambridge University Press.

[11] L. Süssenbach, K. Pitsch, I. Berger, N. Riether and F. Kummert (2012): "Can you answer questions, Flobi?" Interactionally defining a robot's competence as a fitness instructor. In: Ro-Man 2012.