# The Impact of the Contingency of Robot Feedback on HRI


Kerstin Fischer
University of Southern Denmark
Sonderborg, Denmark
kerstin@sdu.dk

Katrin Lohan
Italian Institute of Technology
Genua, Italy
Katrin.Lohan@iit.it

Joe Saunders & Chrystopher Nehaniv
University of Hertfordshire, Hatfield
{j.1.saunders,C.L.Nehaniv}@herts.ac.uk

Britta Wrede & Katharina Rohlfing
Bielefeld University
wrede,@techfak.uni-bielefeld.de
kjr@uni-bielefeld.de



***Abstract*—In this paper, we investigate the impact the contingency of robot feedback may have on the quality of verbal human-robot interaction. In order to assess not only what the effects are but also what they are caused by, we carried out experiments in which naïve participants instructed the humanoid robot iCub on a set of shapes and on a stacking task in two conditions, once with socially contingent, nonverbal feedback implemented in response to different gaze and demonstrating behaviors of the human tutor, and once with non-contingent, saliency-based feedback. The results of the analysis of participants' linguistic behaviors in the two conditions show that contingency has an impact on the complexity and the pre-structuring of the task for the robot, i.e. on the participants' tutoring behaviors. Contingency thus plays a considerable role for learning by demonstration.**



***Keywords- robot as social actor, robot response, robot feedback, timing, contingency, learning by demonstration***


## I. Introduction

In order for robot behavior to count as a 'response' to the behavior of a human user, it has to be perceived as being in a certain temporal, non-accidental relationship to the user's behavior, i.e. it has to be contingent upon the user's behavior (e.g. Watson [46]). In conversation between humans, contingency contributes considerably to joint sense making processes between participants and is an essential feature of human interaction (Hutchby and Wooffitt [20]; Schegloff [37]).

An important aspect of contingency is the temporal proximity between the user's behavior and the robot's response. In natural conversation between humans, communication partners tend to respond to each others' linguistic behaviors in a time frame of 200-300 msecs, with some cultural and stylistic differences observable (cf. Sidnell and Enfield [30]). Furthermore, adjustments have been observed depending on the particular communication partner, for instance, in interactions between adults and children (e.g. Filipi [12]). While in conversation between humans temporal proximity is only partly responsible for the perception of responsivity, it is an important part, and failure to respond in a timely fashion in human conversation is accountable, i.e. in need of explanation (e.g. Levinson 1983: 320 [25]).

Besides the timing of the response, also its content plays a role (cf. Bavelas et al. [2]); in particular, what behaviors are synchronized with what others may provide helpful cues to the interlocutor. For instance, contingency is an important factor in the tutoring strategies of parents when talking about actions and presenting objects to their children. In comparison to action demonstrations performed towards another adult, demonstrations performed towards children are modified such that they exhibit greater contingency: The movements are performed in a tight temporal synchrony with the speech (Gogate et al. [19]) and are shorter, which results in less roundness and more pauses between the individual segments (Brand et al. [3]; Rohlfing et al. [35]). It seems that young infants learn word-object relations within a tightly coupled interaction between infants' perception, joint attention and specific properties of caregivers' naming (Matatyaho and Gogate 2008, p. 172 [28]).

That contingency is adjusted in child-directed speech suggests that it plays a facilitative role in the communication with young children and possibly even in language acquisition; for instance, Fernald and Mazzie [11] demonstrate that when parents are speaking to their children, they synchronize prosodic and content cues, which informs children on the most important words in an utterance and thus helps them segment the speech stream and learn word meanings.

So while contingent response plays a crucial role in interactions between humans, and especially in tutoring situations such as parent-child interactions, what is its role in HRI? Much research in HRI has argued that robot responses should be 'appropriate' or 'timely' (e.g. Breazeal [4]; Steinfeld [41]), yet so far it remains open what 'appropriate' or 'timely' really mean, what exactly the effects of contingency are and, most importantly, what these effects are caused by. In this study, we therefore address to what extent the contingency of a robot's nonverbal responses to the human tutor's actions influences these users' understanding of the human-robot interaction


*This research was partially funded by the European Union in the framework of the* ITALK *project under grant number 214668.*

situation and, concomitantly, their tutoring behavior. We thus address not only what kinds of interactional functions contingency has but also how these effects come about, i.e. what inferences users make about the robot based on its contingent behavior.

## II. Related Work

Yoshikawa et al. (2007) examine the effect of different response latencies of mirroring of blinking on people's perception of being looked at by a humanoid robot. The authors find that people felt most strongly that the robot looked them in their faces when it mirrored their blinking with a 500 msec delay. In contrast, both an immediate and a much delayed (2 secs) reaction created the effect significantly less often. Thus, the timing of the mirroring behavior had a considerable impact on the participants' understanding of the quality of the interaction.

Another relevant investigation is Sidner et al. [39] which addresses the effects of contingent robot gestures in response to human action. In this study, a penguin-shaped robot introduced a novel object to a participant and either used participant-directed head movements and gestures or not. The results of the study show that human communication partners judge a robot producing contingent, non-verbal feedback to be more reliable and its movements more appropriate, and they interact longer with the robot and attend to it more.

Kose-Bagci et al. [24] experiment with different contingency models in non-verbal turn-taking interactions with a humanoid and find that human participants preferred interactions in which a humanoid robot's turn-taking behavior was driven by a contingency model whose temporal dynamics was the closest to 'natural' human-human conversations. Participants in this condition furthermore interacted more and longer with the robot.

A study demonstrating the considerable effect of the contingency between different modalities of the robot's own behaviors is Yamazaki et al. [47]. In this study, the authors compare a systematic condition in which the robot's head movements were timed to co-occur with the transition relevance places of the robot's speech, i.e. points of possible completion (cf. Sacks et al. [36]), with an unsystematic condition in which head turns were not coordinated with the speech. They find that in the systematic condition, people attend to the robot's transition relevance places in their head turns towards the robot and show more engagement by means of nodding than in the unsystematic condition.

Another relevant study is Pitsch et al. [34] in which a robot reacts contingently to visitors in a museum. In particular, the robot detects whether a visitor is looking at it and if so, begins to speak. If the visitor then looks away, the robot pauses briefly and then restarts the utterance. Among humans, this strategy has been found to be efficient to get the partner's full attention, and the authors show that the robot's restarts have a similar effect. Moreover, visitors who heard the restarted utterances stayed longer and reacted significantly more often politely and in similar ways as to other humans to the robot's closing statement. Thus, the robot's contingent display of attention to visitors' attentional states at the beginning of the interaction has a long lasting effect over the course of the entire interaction.

In our own previous work, we have suggested, based on qualitative analyses, that contingent robot behaviors may lead people to increased tutoring behavior (Lohan et al. [26, 27]). That is, people adjusted their action demonstrations to the contingent robot more than to the non-contingent one, resulting in human-robot tutoring interactions that exhibit adjustments similar to those observable in adult-child interaction.

To sum up, previous work suggests that contingency in HRI may have a considerable impact; what this effect may be caused by, and what the interactional consequences really are, is however still open. For instance, while the study by Pitsch et al. [34] has shown that prerecorded monologues by a robot may be perceived as more personal and social if they are delivered depending on the visitor's attentional state, this effect may be due to the fact that it makes visitors realize that the robot perceives them at all; for instance, Nourbakhsh et al. (2003: 3639 [32]) write that 'the single most successful way for a robot to attract human interest is for the robot to demonstrate awareness of human presence.' From this perspective, the contingent response could just be an indicator that the robot is really perceiving the human. On the other hand, the results from the other studies discussed above, that people interact more, longer and are more engaged because of contingent feedback, leave completely unanswered why people react to contingent robot feedback in this way and what the effect is caused by. Thus, why people interacting with a robot that responds 'timely' to a human user's behavior should want to interact with it longer than if the robot is not that 'timely' constitutes a riddle that has no intuitive solution. In the current study, we therefore use a method that allows us to study users' reactions to contingent and non-contingent robot feedback quantitatively while informing us at the same time on the functional differences in people's responses to the different types of robot feedback. In particular, we analyze participants' verbal behavior in the interaction with a contingent and a non-contingent robot as a means to identify their mental models that inform the ways they interact with the robot.

## III. OPERATIONALIZING CONTINGENCY IN THE ROBOT

While robots' behavior may comprise many different modalities, we focus here on how the robot's nonverbal behavior can be coordinated to the human users' nonverbal behavior in terms of contingency, i.e. the production of timely, relevant responses to the human tutor's actions.

In the operationalization of contingency in the system employed in this investigation, contingency is calculated based on temporal co-occurrence of visually detected ostensive signals of human and robot behavior (see Figure 1). In order to model timing, we oriented at prototypical tutoring interactions, namely interactions between mothers and their children. For instance, Keller et al. [23] find that in face-to-face interactions, mothers respond to infants with contingencies within intervals shorter than one second, using a sampling interval of two milliseconds and Watson's [46] method of contingency analysis, across multiple communication modalities. This finding correlates with the findings presented by Stern [42] concerning gaze and head orientation, coded on 16 mm film, 24 frames per second. Likewise, Cohn and Beebe [9] report that most mothers and infants respond to each other with contingencies of less than half a second, using a sampling rate of 1/12 seconds. Van Egeren et al. [45] show that while in a human communication the sender of an ostensive signal usually expects a response within 200-300 msecs, contingent behavior between mothers and infants can be organized within a three second window (they used a sampling interval of one second and an odds-ratio method of contingency analysis). Contingent behavior thus seems to occur within a time frame of 200 - 3000 msecs.

In accordance with these findings, robot feedback in our model occurs within a window of 200-1000 msecs, however, starting no later than 300msecs. With respect to the types of reactions modeled, for the tutoring situations under consideration, the following behaviors of the robot constitute contingent responses to the tutors' behaviors:

- Reaction Pattern 1 (RP-1): the system detects 'participant-gazes-at-elsewhere' and reacts by gazing at random locations and by showing a neutral face,
- Reaction Pattern 2 (RP-2): the system detects 'participant-gazes-at-object' and reacts by directing its gaze at the object and by smiling,
- Reaction Pattern 3 (RP-3): the system detects 'participant-gazes-at-robot's-face' and reacts by directing its gaze to the co-participant and by smiling,
- Reaction Pattern 4 (RP-4): the system detects 'participant-presents-an-object' and reacts by performing a pointing gesture towards the detected location of the demonstration.

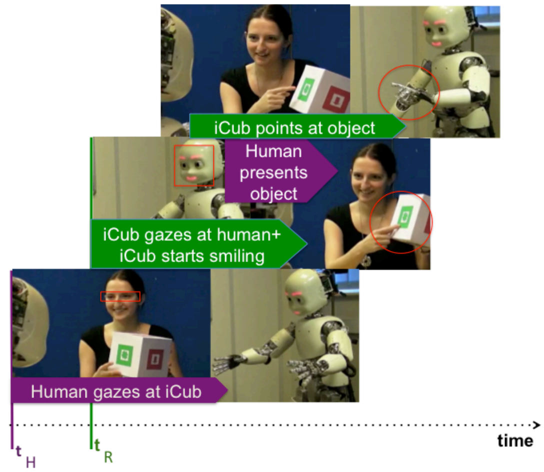


*Figure 1: The implementation controls the robot's interaction with the human tutor by means of responsive behavior as a feedback strategy. The difference between tH and tR is < 300 msecs.*

The implementation monitors participants' hand trajectories with respect to a specific trajectory class, namely presenting behavior. As presenting behavior we define a behavior by the human tutor during which the human tutor moves an object by hand towards the robot and reaches a pre-defined minimal distance. The robot responds to the presenting gesture by trying to point at the object. In principle, the Kinect tracking device used allows easy extension of the system to further trajectory classes, yet currently, object trajectories are calculated if the tutor is looking in the direction of the object. The object is tracked by means of an ARToolkitmarker. The ARToolkit system [1] returns a 3D location of the marker.

The classification of the tutor's eye gaze is obtained by geometrical calculations, resulting from locating the intersection point between gaze direction and the object plane or face plane of the robot. In other words, the eye gaze module detects whether the tutor is looking towards the object, to the face of the learner or elsewhere (see Figure 1). Figure 2 illustrates the interaction between the components of the computational model.

The structure of the robot system, implemented in Java, is summarized in Figure 2. The system is structured into three components: a module providing the robot's responsive behavior, which controls the feedback strategies of the robot, a module analyzing the tutor's behavior, and the contingency detection, which measures online the contingency values of the interaction between the robot and the human.

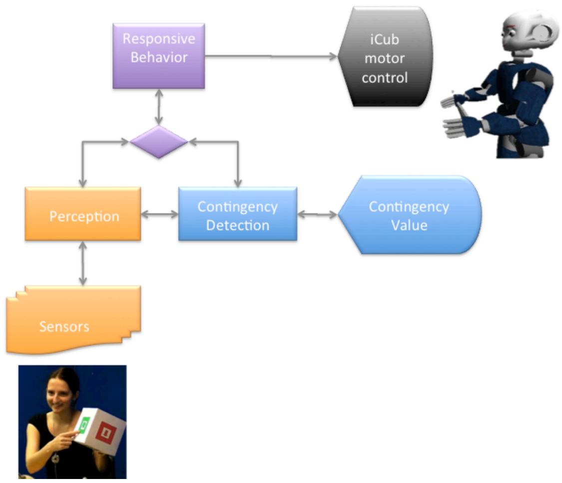


*Figure 2: The structure of the robot system*

The information about whether the tutor is gazing at the robot, at the object or elsewhere and whether the tutor is presenting the object to the robot by holding it at a certain distance towards the robot or not is fed into the contingency module. In addition to information about the human tutor, the contingency module captures also the ongoing behavior of the robot. For the measurement of contingency, thus both interaction partners are taken into account (see Figure 2). Contingency is measured by two variables, the necessity and the sufficiency index. According to Watson, the necessity index describes the forward probability of a consequence given a (hypothesized) cause. From the robot's perspective, the necessity index refers to the probability that the subject's gaze is on a certain object, given that the robot has previously been looking at this object. The sufficiency index measures whether there are also other sources influencing the subject's gazing behavior. For instance, given that the subject's gaze is towards the object, the probability that the robot has previously been looking at this object is calculated. In this interaction, the necessity and the sufficiency indices for the subject's behavior are 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 a product of these two variables. Note that the value for contingency lies thus between 0% and 100%, where 100% means perfect contingency (like for example in a mirror reflection) and 0% no contingency at all. The sufficiency measure in our set-up rises if the tutor is looking at the robot or at the object or if the tutor is presenting an object to the robot. The sufficiency measure decreases when the tutor is looking elsewhere or is not presenting anything. In our scenario, the necessity measure is computed on the robot's behavior and represents the responding behavior of the robot: The measure rises when the robot is looking at the tutor or the object and when it points at an object. It falls when the robot is looking elsewhere or is not showing presentational behavior. The whole calculation is event driven such that each classified behavior is understood as one event. The iCub robot is connected via YARP (Metta et al. [30]) with the system for storing and exchanging data.

## IV. Empirical Study

In the empirical study, we compare human tutors' behavior in two conditions (between subjects): In the contingent condition, the robot responds to the human tutor in socially-contingent ways as described above. In the non-contingent condition, the robot's behavioral inventory is the same as in the first condition, making use of the same set of robot behaviors, yet driven by a saliency model. The non-contingent behavior of the robot is based on the nearest object that the robot would track, and three different arm configurations were executed randomly depending on which side of the robot the object was located; for instance, if the object was on the robot's left, the robot pointed at the object using its left arm. Thus, in both conditions, the robot uses similar behaviors and interacts with its environment; however, in the contingent condition, it responds to social cues while in the non-contingent condition, it responds to features of the object under consideration.

### *A. Participants*

The human-robot interactions were elicited at the University of Hertfordshire in February 2011 and comprise two sets of interactions in two comparable conditions. In both conditions, there were twelve naïve participants who interacted with the robot. In the first condition, three of the twelve participants were female, in the second condition, five of the twelve participants were female. Participants' ages ranged between 21 and 69 (Av: 30, SD: 12). Participants were recruited from university administration staff or were PhD students from various fields and unfamiliar with robotics research.

Participants' instructions were to teach the robot about a set of colored shapes presented on three blocks of different size and then subsequently instruct the robot on how to stack the blocks in size order with the smallest block at the top. While the participants used both verbal and nonverbal instructions, the robot only reacted nonverbally. In the first condition, this behavior was limited to the robot tracking the objects using head and eye movements. In the second condition, the contingency module described above was used.

### *B. Robot*

In both conditions, the robotic participant was the humanoid robot iCub (iCub [21]; Metta et al. [29]).

### *C. Method*

To identify the effects of the amount of contingency of the robot's behavior on the interaction, we investigate in detail users' linguistic behavior in the two conditions. The methodology permits the identification of different behaviors as results of different degrees of contingency in the robot's behavior. The procedure consists of three steps: First, we elicited data in two controlled experimental human-robot interaction scenarios; second, we carried out a quantitative analysis of the linguistic features occurring; third, we apply a qualitative analysis of the functions the linguistic choices users make fulfill in the respective data set (for further details on the method, see Fischer et al. [18]).

Analyses of users' utterances constitute a useful methodology since users' linguistic choices in interaction are correlated with their understandings of these robots (Fischer [16]; Fischer, Lohan and Foth [18]). Given that speakers design their utterances so that they are well suited for the particular communication partner in the current situation (Sacks et al. [36]; Fischer [16]), investigating the linguistic choices speakers make can inform us about what users think about the robot they are interacting with, for instance, what they expect their artificial communication partner to have problems with and what they consider it to be good at. Thus, the association of particular linguistic features with their functions in interaction provides us with qualitative data on speakers' mental models of their artificial communication partners. The linguistic analysis furthermore allows us to understand the effects of robots' behavior since it provides not only objectively identifiable quantitative differences between conditions, but also qualitative measures with respect to the communicative functions to which users attend in the interaction. That is, because people make their linguistic choices with a particular communication partner in mind, the functional characteristics of their utterances allow us to infer what effects the robot's behavior in the two conditions have on participants' conceptualizations of the robot and consequently on the quality of the HRI interaction.

The linguistic analyses thus provide evidence not only on whether socially-contingent robot behavior improves human-robot interactions, but if so, also on why this is the case since tutors' linguistic choices are directly related to certain speech functions, which inform us as analysts on tutors' understandings of the affordances of the respective situation.

### *D. Procedure*

Participants in both conditions were asked to teach a set of shapes to the robot. Their utterances were transcribed and linguistically analyzed by semi-automatically extracting certain linguistic properties from participants' utterances for which linguistic research has established a set of communicative functions, which can be informative of what users suspect to be at issue in a given situation. Thus, no extra data encoding is necessary, just the extraction of the occurrences of certain linguistic features in users' utterances. These behavioral measures were chosen since unlike post-experimental questionnaires and other indirect methods, they provide online indicators of users' understandings of the affordances of the communicative situation.

### *E. Data Analysis*

The set of features used covers both complexity levels and interpersonal aspects of interactions. The features used here have been employed in numerous previous studies (e.g. Fischer [14-16]; Fischer et al. [17, 18]), where they have been shown to be reliable indicators for the communicative functions to which people attend in interaction; they thus allow us to infer users' understandings of the affordances of the respective interaction.

The set of linguistic features used comprises:

- verbosity, in particular the number of turns and words used in the interaction with the robot; these measures also serve to establish a baseline for the other measures;

- structuring cues, in particular the numbers of instances of local structuring items, such as *now*;

- attention-getting, for instance, by means of imperatives like *look* or the robot's name;

- grounding, in particular the number of items with which users refer to entities that can be assumed to be sufficiently grounded (cf. Clark & Schaefer [8]); i.e. linguistic means to refer to items from the discourse record are, for instance, *again* or *other*.

- diversity, in particular the number of different words used;

- interpersonal relationship, in particular the kinds of pronouns used to refer to the participants, for instance, *I* versus *you* versus *we*; furthermore, evidence on the supposed interpersonal relationship are also the amount of involvement of the communication partner by means of tag questions, questions or, in comparison, statements;

- complexity, in particular whether expository sentences are used in order to reduce the cognitive complexity of an utterance by presenting an object first before something is asserted about it. In the current dialogs, this is usually done by expository utterances like, *this is an X, and it has property Y*. In addition, the number of utterances per turn serves as an indicator for complexity, as well as the Mean Length of Utterance (MLU), i.e. the number of words per sentence.

The feature set used is intended to cover a broad range of respects in which people adjust their speech to particular communication partners, such as children (e.g. Snow [40]), foreigners (Ferguson [10]), or robots (Fischer [13]). Thus, the point of the analysis is not that speech to a particular communication partner should differ in all respects from other kinds of speech; instead, what is relevant for the analysis is in what respects linguistic behaviors in the two conditions compared differ, since this reveals to which communicative functions people attend (differently in the two conditions) and what they consider relevant for a particular communication partner.

## V. RESULTS

The linguistic analyses show that users' linguistic behaviors in the two conditions are significantly different with respect to several of the linguistic features investigated; however, the differences concern only specific kinds of linguistic features.

First, we find consistent attempts at structuring the information to be learnt for the contingent robot, but less so for the non-contingent robot; in particular, users produce significantly more utterances that introduce a shape first before something is asserted about it, thus decomposing the complex task into smaller parts, i.e. there are significantly more instances of expository utterances.

Furthermore, there are significantly more structuring cues and higher order structures in speech to the contingent robot. In particular, while the number of turns is comparable across conditions, the numbers of both questions and statements are higher in the contingent condition, which suggests that people use more, shorter utterances within the same number of turns for the contingent robot. At the same time, users produce fewer different words for the contingent robot. Thus, input to the contingent learner consists of longer turns with more, shorter and simpler utterances.

There are furthermore tendencies for users in the interaction with the contingent robot to understanding the situation as a joint project (Clark [7]) since they tend to use more *let's* in this condition. However, one might also expect more instances of the pronoun *we* by means of which users refer to themselves and the robot together; these data do not reach significance. Instead, we observe fewer instances of *I* in the interaction with the contingent robot ($p < .08$). On the whole, however, there are only few differences with respect to the interpersonal relationship between human and robot in the two conditions.

That the amounts of tag questions are significantly lower in the contingent interactions, yet that the numbers of questions tend to be higher could be an indicator that participants do not impose any understanding on the robot but treat it as a serious communication partner; since tag questions basically ask the co-participant to simply agree to the statement made, it does not leave much for the partner to do, in contrast to a real question. However, tag questions also have interactive and attention-managing functions, and thus the interpretation of this linguistic difference is not without problems.

What clearly remains the same across conditions are the users' attempts at getting the robot's attention; thus, instances of the robot's name, instances of attention-getting *look* are not significantly different across conditions.

Table 1 presents the ANOVA results for the two conditions.

*Table 1: ANOVA results for the contingent versus non-contingent robot (t = $p < .10$; * = $p < .05$; ** = $p < .01$; *** $p < .001$)*

| | **Mean non-contingent** | ***sd*** | **Mean contingent** | ***sd*** | **F(1,44)** | **p** |
|---|---|---|---|---|---|---|
| verbosity: turns | 118.5 | 45.6 | 122.7 | 32.2 | 0.13818 | |
| verbosity: words | 481.9 | 203.6 | 599.6 | 235.5 | 3.30534 | t |
| diversity: different words | 0.21 | 0.06 | 0.18 | 0.03 | 6.57359 | * |
| complexity: MLU | 3.86 | 1.12 | 4.82 | 1.30 | 7.21429 | *** |
| complexity: expository | 0.06 | 0.03 | 0.11 | 0.08 | 9.25134 | ** |
| structuring: *now* | 0.01 | 0.01 | 0.01 | 0.01 | 4.48447 | * |
| grounding: *other* | 0.02 | 0.03 | 0.01 | 0.01 | 3.62421 | t |
| grounding*: again* | 0.02 | 0.02 | 0.03 | 0.02 | 2.91095 | t |
| interpersonal: *I* | 0.02 | 0.03 | 0.01 | 0.01 | 3.22947 | t |
| interpersonal: *we* | 0.06 | 0.00 | 0.01 | 0.02 | 0.34817 | |
| interpersonal: *let's* | 0.00 | 0.00 | 0.01 | 0.02 | 3.57125 | t |
| interpersonal: *you* | 0.04 | 0.04 | 0.05 | 0.07 | 0.17627 | |
| interpersonal: feedback | 0.02 | 0.02 | 0.01 | 0.01 | 2.04759 | |
| interpersonal: tags | 0.00 | 0.01 | 0.00 | 0.00 | 4.35391 | * |
| interpersonal: questions | 0.01 | 0.02 | 0.02 | 0.02 | 3.08320 | t |
| interpersonal: statements | 0.83 | 0.18 | 0.93 | 0.03 | 6.53347 | * |
| attention: robot's name | 0.00 | 0.00 | 0.01 | 0.03 | 0.58112 | |
| attention: *look* | 0.00 | 0.01 | 0.03 | 0.01 | 0.29942 | |

To sum up, these results show that people reduce the complexity of their utterances more when speaking to the contingent than to the non-contingent robot, and that they structure their instructions more and decompose complex issues into smaller units. At the same time, they trust the contingent robot more to learn from the interaction, as evidenced by the marginally higher numbers of instances of *other* and *again*. On the other hand, contingency seems to have no effect on attention-getting functions in interaction, and only a small impact on the interpersonal relationship between robot and tutor.

## VI. DISCUSSION

The data show that the contingency of robot behaviors has several significant effects on the interaction. Regarding the structuring of information, we found that users provide the contingent robot with local (*now*) and global (*again, other*) clues to discourse structure; especially the latter presuppose considerable cognitive capabilities and the collaborative achievement of a common ground. At the same time, we find attempts at simplification (reduced diversity, expository strategies), which results in longer turns consisting of fewer, shorter utterances. Thus, users present the information in smaller chunks for the contingent robot. This is in line with a previous, qualitative investigation of the same data (Lohan et al. [27]), in which a sequential micro-analysis showed how users in the contingent condition adjust their utterances to the robot's suspected needs and capabilities. The current findings are also in line with results from Fischer et al. [17], in which people did not adjust their verbal behavior to a simulated robot designed to look like a young child, as the comparison with child-directed speech in the same scenario revealed. However, people did adjust their gestures to the robot's eye gaze since this provided them with online contingent feedback on what the robot was seemingly able to understand.

The results from the current study furthermore point into the same direction as results from other asymmetrical interactions, such as interactions with children. In particular, the findings of the current study are comparable to the results of a study by Murray and Trevarthen [31], in which the contingency of the infants' responses was at issue. In this study, the authors recorded eight mothers in a video-mediated interaction with their two-month olds during which mothers believed that they were seeing their infants live the whole time; yet only in half of the interactions the live image of their infant was transferred. The authors find significant differences in mothers' behavior depending on whether mothers saw their infants live or as replay. The linguistic analysis shows that mothers asked more questions, repeated their utterances more often and produced fewer negative statements and fewer declaratives if their infant was not behaving contingently (Murray and Trevarthen 1986: 23 [31]). The linguistic features affected in these interactions thus also concern the complexity of the mothers' utterances; these results strengthen the observations made here, that contingent robot feedback has an impact on speakers' assumptions about the communication partner's competences.

The findings from the current study consequently demonstrate that the effects of contingent feedback concern most importantly users' understanding of the robot's competence. We can conclude that one reason for the considerable impact of contingency is that people interpret it as online indicator of the robot's understanding of the interaction. From this perspective, a contingent robot response is taken by participants as reliable information about the robot's processing capabilities in general and about its processing of the current state of talk in particular; contingent robot feedback thus reduces users' uncertainty with respect to their unfamiliar, artificial communication partner by virtue of being produced at the 'right' moment, where the identification of the 'right' moment leads to inferences about the robot's general cognitive capabilities.

## VII. CONCLUSION

The analysis confirms that the amount of social contingency in robots' behavior has a considerable impact on users' expectations and their subsequent tutoring behavior. In addition, the results of the current investigation allow us to understand what this impact is caused by. In particular, the considerable impact of contingent robot behavior was shown to be due to the fact that people interpret contingent robot feedback as direct, unmediated and trustworthy indicators of the robot's understanding of the current state of talk. This interpretation feeds into people's mental models of their artificial communication partner, such that they build up a coherent image of the robot's capabilities, strengths and weaknesses.

## VIII. DESIGN IMPLICATIONS

The results regarding the ways information is structured and presented to the robot are likely to have an impact on the success of socially guided learning (see Thomaz and Breazeal [43], Thomaz and Cakmak [44], Cakmak et al. [5, 6]). As Thomaz and Cakmak [44], for instance, demonstrate, naïve users structure information for robots intuitively in ways that improve learning from demonstration. If the ways in which this is done are influenced by the amounts of contingency in the behavior exhibited by the robot, contingency can be inferred to have a considerable influence on the success of learning by demonstration.

## ACKNOWLEDGMENTS

We thank Dr. Hagen Lehmann of the University of Hertfordshire for expert assistance with the experiments and Dr. Sandra Folgate for the transcription of the data.

## REFERENCES

[1] ARToolKit; http://www.hitl.washington.edu/artoolkit/; last accessed: 11/03/12

[2] J. Bavelas, Coates, L. and Johnson, T., Listener Responses as a Collaborative Process: The Role of Gaze. *International Communication Association*, pp. 566-580, 2002.

[3] R.J. Brand, Baldwin, D. A. and Ashburn, L. A., Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science* 5: 72–83, 2002.

[4] C. Breazeal, Toward Sociable Robots. *Robotics and Autonomous Systems* 42: 167-175, 2003.

[5] M. Cakmak, and Thomaz, A., Optimality of Human Teachers for Robot Learners. *Proceedings of the 9th IEEE International Conference on Development and Learning,* pp. 64-69, 2010.

[6] M. Cakmak, N. DePalma, A.L. Thomaz and R. Arriaga, Effects of Social Exploration Mechanisms on Robot Learning. *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication* (Ro-MAN09), 2009.

[7] H.H. Clark, *Using Language*. Cambridge University Press, 1996.

[8] H.H. Clark and Schaefer, E.F., Contributing to Discourse. *Cognitive Science* 13, 259-294, 1989.

[9] J. Cohn and Beebe, B., Sampling interval affects time-series regression estimates of mother-infant influence. *Infant Behavior and Development,* Abstracts Issue, 13: 317, 1990

[10] C.A. Ferguson, Towards a Characterization of English Foreigner Talk. *Anthropological Linguistics* 17, 1, 1-14, 1975.

[11] A. Fernald and Mazzie, C., Prosody and Focus in Speech to Infants and Adults. *Journal of Developmental Psychology* 27: 209-221, 1991.

[12] A. Filipi, *Toddler and Parent Interaction*. Amsterdam, John Benjamins, 2009.

[13] K. Fischer, *What Computer Talk Is and Isn't: Human-Computer Conversation as Intercultural Communication*. Saarbrücken: AQ, 2006.

[14] K. Fischer, Interpersonal Variation in Understanding Robots as Social Actors. *HRI'11*, Lausanne, 2011.

[15] K. Fischer, How people talk with robots – Designing Dialog to Reduce User Uncertainty. *AI Magazine*, 32(4): 31-38, 2011.

[16] K. Fischer, Recipient Design, Alignment, Interaction: The Role of the Addressee in so-called 'Simplified Registers'. Habilitation Thesis, University of Bremen, 2011.

[17] K. Fischer, Foth, K., Rohlfing, K. and B. Wrede, Mindful Tutors: Linguistic Choice and Action Demonstration in Speech to Infants and a Simulated Robot. *Interaction Studies* 12, 1: 134-161, 2011.

[18] K. Fischer, K., Lohan, K. and K. Foth, Levels of embodiment: Linguistic analyses of factors influencing HRI In *Proceedings of HRI'12*, Boston, MA., March 2012.

[19] L.J. Gogate, Bahrick, L. E. and Watson, J., A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development* 71: 878–894, 2010.

[20] I. Hutchby and Wooffitt R., *Conversation Analysis*. Polity Press, Cambridge, UK, 1998.

[21] iCub, "RobotCub – An Open Framework for Research in Embodied Cognition," http://www.robotcub.org/, 2004.

[22] iCubForwardKinematics 2008, http://eris.liralab.it/wiki/ICubForwardKinematics; last accessed: 13/03/12.

[23] H. Keller, A. Lohaus, S. Völker, M. Cappenberg, and A. Chasiotis, Temporal contingency as an independent component of parenting behavior. *Child Development,* 70(2):474–485, 1999.

[24] H. Kose-Bagci, Dautenhahn, K., Syrdal, D.S. and Nehaniv, C.L., Drum-mate: Interaction dynamics and gestures in human-humanoid drumming experiments, *Connection Science*, 22(2):103-134, 2010.

[25] S. Levinson, *Pragmatics*. Cambridge University Press, 1983.

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

[27] K. Lohan, K. Pitsch, K.J. Rohlfing, K. Fischer, J. Saunders, H. Lehmann, C. Nehaniv and B. Wrede, Contingency allows the robot to spot the tutor and to learn from interaction. In*: Proceedings of ICDL* 2011, Frankfurt.

[28] 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.

[29] G. Metta, Sandini, G. Vernon, D. Natale, L. & F. Nori, The iCub humanoid robot: an open platform for research in embodied cognition. *PerMIS: Performance Metrics for Intelligent Systems Workshop.* August 2008, pp. 19–21, 2008.

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

[31] L. Murray and Trevarthen, C., The Infant's Role in Mother-Infant Communications. *Journal of Child Language* 13, 15-29, 1986.

[32] I.R. Nourbakhsh, Kunz, C. and Willeke, T., The Mobot Museum Robot Installations: A Five Year Experiment. *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, Nevada, October 2003, pp. 3636-3641.

[33] K. Pitsch, Kuzuoka, H., Suzuki, Y., Süssenbach, L., Luff, P & Heath, C., 'The first five seconds': Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. *The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan, Sept. 27-Oct 2, 2009, pp. 985-991.

[34] K. Pitsch, Lohan, K., Rohlfing, K., Saunders, J., Nehaniv, C. and Wrede, B., Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human-robot-interaction. *RoMan* 2012.

[35] K.J. Rohlfing, Fritsch, J., Wrede, B. and Jungmann, T., How can multimodal cues from child-directed interaction reduce learning complexity in robots? *Advanced Robotics* 20(10): 1183-1199, 2006.

[36] H. Sacks, Schegloff, E.A. and Jefferson, G., A Simplest Systematics for the Organization of Turn Taking for Conversation. *Language* 50: 696–735, 1974.

[37] E.A. Schegloff, Issues of Relevance for Discourse Analysis: Contingency in Action, Interaction and Co-Participant Context, in E. H. Hovy and D. Scott (eds.), Computational and Conversational Discourse: Burning Issues - An Interdisciplinary Account. Heidelberg: Springer, pp. 3-38, 1996.

[38] J. Sidnell and Enfield, N. J., Language diversity and social action: A third locus of linguistic relativity. *Current Anthropology* 53, 302-333, 2012.

[39] C.L. Sidner Lee, C.L., Kidd, C., Lesh, N. and Rich, C., Explorations in Engagement for Humans and Robots. *Artificial Intelligence* 166, 1-2: 140-164, 2006.

[40] C.E. Snow, Beginning from baby talk: Twenty years of research on input and interaction. In: Gallaway, C. & Richards, B.J., eds. Input and Interaction in Language Acquisition, 3-12. Cambridge: Cambridge University Press, 1994.

[41] A.Steinfeld, T.Fong, D.Kaber, M.Lewis, J.Scholtz, A.Schultz and M.Goodrich. Common Metrics for Human-Robot Interaction. *HRI '06*, Salt Lake City, 2006.

[42] Stern, D.N. 1971. A micro-analysis of mother-infant interaction. behavior regulating social contact between a mother and her 3 1/2 month-old twins. *Journal of the American Academy of Child Psychiatry,* 10(3): 501–517.

[43] A.L. Thomaz and Breazeal, C., Teachable robots: Understanding human teaching behaviour to build more effective robot. *Artificial Intelligence* 172: 716–737, 2008.

[44] A.L. Thomaz and Cakmak, M., Learning about objects with human teachers. *HRI 2009*, pp. 15-22, 2009.

[45] L.A. Van Egeren, Barratt, M.S. and Roach, M., Mother-infant responsiveness: Timing, mutual regulation, and interactional context. Developmental Psychology, 37, 684-697, 2001.

[46] J.S. Watson, Contingency perception in early social development. *Social Perception in Infants,* pp.157–176, 1985.

[47] A. Yamazaki, Yamazaki, K., Kuno, Y. Burdelski, M. Kawashima, M. and Kuzuoka, H., Precision Timing in Human-Robot Interaction: Coordination of Head Movement and Utterance. *CHI 2008*, April 5-10, 2008, Florence, 2008

[48] Y. Yoshikawa, Shinozawa, K. and Ishiguro, H., Social reflex hypothesis on blinking interaction. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, pp. 725-730, 2007.

[49] P. Zukow-Goldring, Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. In: *Early Development and Parenting* 5: 195-211, 1996.