Levels of Embodiment: Linguistic Analyses of Factors Influencing HRI

Kerstin Fischer
University of Southern Denmark
IFKI, Alsion 2
DK-6400 Sonderborg
0045-6550-1220
kerstin@sitkom.sdu.dk

Katrin Lohan
Bielefeld University
Universitätsstr. 23-25
D-33615 Bielefeld
klohan@uni-bielefeld.de

Kilian Foth
University of Hamburg
Department of Informatics
Vogt-Koelln-Str. 30
D-22527 Hamburg
foth@informatik.uni-hamburg.de

ABSTRACT

In this paper, we investigate the role of physical embodiment of a robot and its degrees of freedom in HRI. Both factors have been suggested to be relevant in definitions of embodiment, and so far we do not understand their effects on the way people interact with robots very well. Linguistic analyses of verbal interactions with robots differing with respect to physical embodiment and degrees of freedom provide a useful methodology to investigate factors conditioning human-robot interaction. Results show that both physical embodiment and degrees of freedom influence interaction, and that the effect of physical embodiment is located in the interpersonal domain, concerning in how far the robot is perceived as an interaction partner, whereas degrees of freedom influence the way users project the suitability of the robot for the current task.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Human Factors

Keywords

Embodiment, robot simulations, degrees of freedom, linguistic analysis, verbal human-robot interaction

1. INTRODUCTION

Previous work has identified considerable problems in defining the notion of embodiment. Ziemke [33], for instance, suggests six different dimensions along which the notion can be defined: Embodiment as structural coupling (between the organism and its environment); historical embodiment (comprising also the history of previous structural couplings); physical embodiment, accounting for the common sense notion of having a physical body; organismoid embodiment, having at least to some extent a similar form and sensorimotor equipment as living organisms; and organismic embodiment, i.e. being a living organism; and finally, social embodiment (concerning the functions of embodiment in

social interaction). These dimensions reflect the different criteria discussed from a broad range of perspectives and in different disciplines, for instance, cognitive science, biology, robotics and cognitive linguistics. In contrast, other authors focus on a minimal definition of embodiment; for instance, Dautenhahn et al. [3] aim at identifying criteria that define the core of the concept. The authors suggest perturbatory bandwidth and structural variability to be central criteria; they define perturbatory bandwidth in terms of the sensitivity of the system to react to, and to act upon, its environment, i.e. in the amount of sensors and the way the information from the sensors is made use of. More sensors and more degrees of freedom thus allow a robot to interact more with its environment and hence the robot will be embodied to a greater degree. Structural variability, on the other hand, refers to the system's ability to adapt to, and change through, the interaction with its environment. Their definition allows the quantification of degrees of embodiment: $DOM_{S,E} = f(x,y,t)$, where the degree of embodiment (DOM) of a system S is defined with respect to an environment E as a function of the vectors x and y and time t. The two vectors represent perturbatory bandwidth and structural variability respectively.

Concerning the effects of these different levels of embodiment, previous work in HRI has mostly focused either on the role of anthropomorphic properties of artificial agents on the one hand, or on the role of physical embodiment on the other; regarding the former, numerous studies have shown that anthropomorphic features of robots trigger social responses from participants (e.g. Sproull et al. [29], Koda & Maes [17], Nass [26]); regarding the latter, most studies suggest that users prefer real robots over simulations and present robots over telepresent ones; thus, physical embodiment seems to play a crucial role. Other kinds and degrees of embodiment have hardly been investigated, and so far it is unclear what exactly the different findings are caused by. Consequently, we still understand little of the effects of different types and degrees of embodiment on the processes involved in human-robot interaction (cf. Kiesler et al. [16]). In this study, we approach aspects of embodiment using a relatively new methodology that has several advantages for the investigation of the effects of certain design decisions on HRI; in particular, we address the role of physical embodiment on the one hand and of different degrees of freedom of the robot on the other. We will show how users' perceptions of systems with different embodiment differ using linguistic analysis which will provide additional information on the influences of certain degrees and kinds of embodiment.

2. RELATED WORK

Related work concerns, for instance, the evaluation of the robot, for instance, with respect to its enjoyability, attributions of intentions and judgments concerning its social presence. These effects are usually assessed using post-experimental questionnaires. In addition, most studies also address one or more behavioral measures which concern, for example, the ability of the robot to persuade the user to do something, to speed up the users' performance, or to facilitate learning.

The results of the different studies do not allow unanimous conclusions, however. While several studies demonstrate an effect for physical embodiment, others do not. For instance, Lohan et al. [20] investigate the effect of physical embodiment on participants' behavior in interaction and show that participants interact more with a real robot than with a simulated one in terms of eye gaze. However, studies focusing on other variables do not find consistent effects of physical embodiment. For instance, Komatsu & Abe [18] compare the effects of a simulated and a physically embodied robot in a setting in which the robots distracted the participants from their current activity. Dependent variables were whether participants followed the robots' invitations, where participants looked and their success rate in the game. There was a tendency for more participants to switch the activity with the physically embodied robot than with the simulated robot, yet there were no effects on looking times or task completion. In a follow-up study, Komatsu & Kuki [19] find manipulations of speaker expectations to influence participants' willingness to follow the simulated and the physical robot' invitations to the same degree. In a further study, Komatsu & Kuki [20] find that expectation manipulation (by means of an text stressing the life-like interactive capabilities of the robot) makes up for the initial disadvantage of the simulation. Similarly, Hoffmann & Krämer [13] find pleasantness of physical versus simulated robots to depend on the task to be solved by the user: In task-based problem solving, the physical robot was evaluated better, whereas in a conversational setting, the screen-based robot was rated higher. The authors furthermore find no differences with respect to social presence between the two conditions, yet their participants felt more control in the scenario with the agent displayed on a screen. Moreover, consistent with the other studies, there were no significant differences in participants' performance in the two tasks depending on the condition. Finally, contrary to the results by Komatsu and colleagues, in their experiment physical embodiment had no effect on the persuasiveness of the respective robot. To sum up, there seem to be effects for physical embodiment, yet they seem to be mediated by other factors.

Bainbridge et al. [1] investigate the role of the physical presence of a robot using a real and a tele-present robot in their experiments. They find in post-experimental questionnaires that participants give generally higher scores to the physically present than to the video-displayed robot, though very few differences in questionnaire items individually reached significance. Regarding the behavioral measures, they find no significant differences between conditions regarding the reciprocation of the robot's greeting, yet participants followed the robot's unusual request to trash some expensive books significantly more often in the situation in which the robot was physically present. Finally, they find more participants to respect the robot's personal space in the physical presence condition than in the video condition. These results are in contrast to the results by Kidd & Breazeal [15] who find no differences between a present robot and one whose image

is displayed on the screen. Instead, they report considerable differences in the perception of physically embodied robots in comparison with an animated character. Thus, they find presence to have no impact, whereas the reality of the agent makes a difference.

Wainer et al. [31] aim to disentangle the effects of physical embodiment and presence in a task-based scenario. Their participants (mostly computer scientists and even roboticists) find the embodied robot to be more appealing and perceptive of the world than the virtual robot, and in a post-experimental questionnaire, they judge the present physically embodied robot to be more enjoyable and helpful than both the virtual and the tele-present robot. There are no significant differences between the simulated and the tele-present robot, and the present, embodied robot scores highest in all comparisons. Likewise, Kiesler et al. [16] compare four different conditions, in which the robot is a) present, b) remote and projected, c) simulated and present and d) simulated, remote and projected. They find physical embodiment to matter such that users spend more time with the physically embodied robot, disclose less and attribute more human behaviors and intentions to it, while presence has no impact on these outcomes. In contrast, presence made a difference on users' behavior (healthy eating) after the experiment and on their judgments of the robots' enjoyability.

Regarding degrees of freedom, there are no conclusive results indicating that the amount of flexibility of a robot should influence the way people interact with the respective system, either. For instance, Lusk & Atkinson [23] investigate degrees of embodiment in pedagogical agents; the dimension they address concerns the amount of movement of the animated figure. The authors compare students' learning behavior in three conditions: a multimedia environment either equipped with a locomoting agent in form of a parrot which uses gaze and gesture to direct attention, the same agent which was however minimally embodied such that it remained static on the screen, and a voice only condition with no animated character. The authors find a higher level of understanding for the static agent compared to the no agent condition and significantly better transfer of the material learned in the animated agent condition than in the no agent condition. The effects are however small, and the behavior of the agent seems to play a rather neglectable role, at least with respect to learning.

To sum up, previous work shows some tendencies regarding possible effects of physical embodiment, realism, and degrees of freedom, yet there is no conclusive picture emerging. Thus, a qualitative methodology targeting users' cognitive representations of their artificial communication partner may be useful to investigate the role of different degrees and kinds of embodiment.

3. METHODOLOGY

Since the methodology proposed is rather unusual in HRI research, we describe it in some detail.
The key principle exploited in this method is the notion of recipient design (Sacks et al. [28]). People have been found to choose the linguistic features of their utterances on the basis of their understanding of the needs of the current communication partner; for instance, in child-directed speech, caregivers adjust their utterance length to the child's knowledge of the words within

that utterance, thus taking considerable knowledge about their communication partner into account (Roy et al. [27]). Similarly, caregivers adjust syntactic, semantic and pragmatic features of their utterances to their children's respective receptive capabilities (Cross et al. [2]). Concerning speech to robots, recent studies also show that users' linguistic choices in interaction are correlated with their understandings of these robots (Fischer [7], Fischer et al. [8]). Given that speakers design their utterances so that they are well suited for the particular communication partner in the current situation, investigating the linguistic choices speakers make can inform us about what users judge the communication partner to have problems with and what they consider them 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 method proposed has a number of advantages:

- it relies on objectively observable behavior and avoids the pitfalls of self-reports;

- it is comprehensive, broad and well suited both for discovery and for hypothesis testing;

- it measures users' behavior in the moment; Fussel et al. [10], for instance, show that ratings of anthropomorphism may differ considerably depending on whether they are elicited during interactions or afterwards (see also Takayama [30]);

- and it is online and can thus be exploited for adapting to the respective user once the relationship between linguistic features and user preferences has been established.

The methodology therefore allows us not only to identify different behaviors as effects of different degrees and types of embodiment, but it provides us also with information on the functional effects of different degrees and types of robot embodiment.

The methodology employed consists of three parts: 1) data elicitation in controlled experimental human-robot interaction scenarios; 2) quantitative analysis of the linguistic features occurring; 3) qualitative analysis of the functions the linguistic choices users make fulfill in the respective data set.

Data elicitation takes place in experimental settings that are controlled concerning all external factors and differ only with respect to the variable investigated. Within the data elicitation sessions, participants are not restricted in what they say or how they behave. They are asked to interact verbally with the robot, irrespective of whether or not the robot produces utterances (yet) itself (in the current scenario this is particularly credible since the robots employed were designed to resemble young children).

The data analysis concerns the linguistic properties of participants' utterances in the interaction with the robot. These properties serve as dependent variables and provide the analyst with evidence for particular ways of understanding the affordances of the respective robot, tasks and situation. Thus, the linguistic choices participants make reveal the mental models they build up of the robot as a communication partner during the interaction.

The data are recorded and subsequently transcribed, that is, an orthographic version of the spoken utterances is created. This work is typically done by assistants who are ignorant about the goals of the study, and transcriptions are usually crosschecked in order to guarantee a trustworthy rendering of the spoken data.

The quantitative data analysis is done half-automatically using computational tools. The qualitative analysis is based on linguistic research on the functions of the linguistic features in question.

4. EMPIRICAL STUDY

The study comprises the elicitation of human-robot interactions in three different conditions that differ concerning the robots involved. In the current study, the independent variable concerns the kind of robot and the degrees of freedom of the robots employed. The different data sets were thus elicited with the same tasks the participants had to fulfill, namely to explain the use of some household objects (e.g. a lamp, a salt carrier, a bell) and some toys (e.g. cups, blocks and a bag with rings) to the respective robot. The dependent variables investigated are users' spontaneous linguistic behaviors in these tasks.

4.1 Robots

The first robot (henceforth Akachan) is simulated such that it is presented on a screen. However, it interacts with its environment by means of eye-gaze, which is driven by a saliency model (Nagai & Rohlfing [25]). This visual attention module detects salient locations in a scene which stand out in terms of color, intensity, changes in brightness and motion. The calculation of the salient point follows the model by Itti et al. [14]. The robot, originally called 'Babyface', was designed to resemble an infant, yet in order not to influence participants, we used the Japanese translation 'Akachan' to refer to it in the experiments. In addition to the saliency-based eye-gaze, the robot blinks randomly with its eye lids and opens its mouth in a random fashion.

The second robot (henceforth iCub I) is the humanoid robot iCub (Metta et al. [24]), which is physically embodied with 53 degrees of freedom, yet to make it comparable to Akachan in Condition 1, it was enabled only to move its eyes. Its gaze behavior was controlled by the same mechanism as Akachan in the first condition.
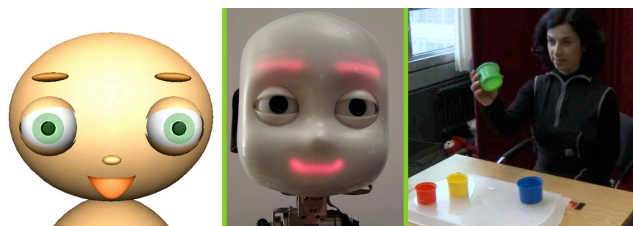

Figure 1: left: Babyface/Akachan; mid: iCub; right: a participant explaining how to stack cups

The third robot (henceforth iCub II) is again iCub, but here the robot is allowed to move its head together with its eyes. Thus, of the 53 degrees of freedom available to the robot, only those controlling the eyes, the head and the neck were enabled. In this way, the robot exhibits exactly the same functionality as the other two robots – it gazes at a salient object or movement, yet it does so by moving both head and eyes. The saliency module was the same as in the other two conditions. The coordination between eye and head movements was modeled based on Lopes et al. [22].

4.2 Participants

In the first condition, 30 participants interacted with Akachan. In the second condition, eight participants interacted with iCub I (which was only moving its eyes), and six participants interacted with iCub II (which was moving its eyes and head). All participants were students or staff from Bielefeld University from various different disciplines. Participants were between 17 and 63 years old and recruited by word-of-mouth.

4.3 Data Elicitation

Data elicitation proceeded in three different conditions: In condition 1, participants were seated at a table across a screen showing the simulated robot Akachan. The experimenter placed objects on the table in front of the participant, which the participant then had to explain to the robot. The robot followed the instructions with its eyes.
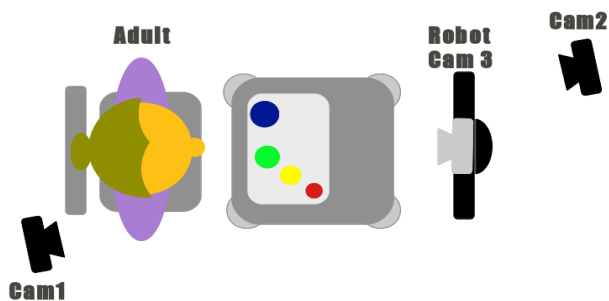

Figure 2: Experimental set-up for Condition 1 using the simulated robot Akachan

In the second condition, participants were seated across the table of the humanoid robot iCub. Again, objects were placed in front of the participant with the request to explain these objects to the robot. As in Condition 1, the robot follows the instructions by means of eye gaze.

The set-up for the third condition is identical to the second condition with the only difference that the robot (iCub II) in this condition moves eyes and head together.
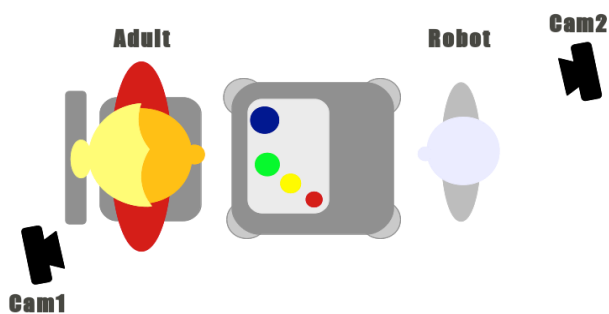

Figure 3: Experimental set-up for Conditions 2 and 3 using the humanoid robot iCub

4.4 Data Analysis

The utterances produced by the participants were manually transcribed and syntactically analysed. The linguistic analysis was carried out using the constraint-based parser described in Foth et al. [9]. This system performs morphological classification and syntactic and referential dependency analysis on the word level and assigns every dependency to one of 35 syntactic classes. The output format allows the quick computation of basic frequency counts such as mean length of utterance (MLU) or category distribution, but also supports searches among inflected words for their stems, or for the syntactic roles of words. The label set employed allows distinctions such as those between subjects, direct objects, and indirect objects, or between active and passive voice, to be retrieved easily. To rule out distortions of the results due to any systematic imperfections of the parsing accuracy, all analyses were fully verified for correctness manually. The linguistic analysis concerns three different factors: verbosity, complexity and interactivity.

Linguistic verbosity
The first general property investigated concerns the amount of speech presented to a communication partner, i.e. linguistic verbosity. The verbosity measures tell us about how much effort speakers spend on each task and how much information they consider suitable or necessary for their communication partner to understand, thus providing indirect information about speakers' recipient design for their respective communication partners. Moreover, the number of different words tells us about the suspected competence level of the communication partner. Thus, to begin with, for each corpus we counted the total number of words for further analyses and the number of different words per speaker in each of the six tasks (diversity) as well as number of utterances per task.

Complexity of utterances
The second measure concerns the complexity of utterances; a very common measure of sentence complexity is the **MLU**, the mean length of utterance. To calculate the MLU, we simply divided the number of words per speaker by the number of utterances by the same speaker. By utterance we understand all turn-constructional units, that is, units consisting of clause complexes, of single clauses, but also smaller units, such as noun, verb or prepositional phrases, answer particles and feedback signals that occur independently (Sacks et al. [28]).
Another measure of complexity, and at the same time a feature revealing the suspected competence of the communication partner, concerns the **concreteness** versus **abstractness** of terms used, for instance, *cup*, *bowl*, or *block* versus *object, container* or *obstacle* (cf. Fischer [5]).
Furthermore, some structures are more complex than others. The **passive**, for instance, is a structure that introduces a perspective in which the patient or undergoer of an action is foregrounded and the agent is backgrounded. The construction is also formally quite complex and thus a useful indicator for assumed competence.
Sentence complexity is also reflected in the number and type of objects used. In particular, we distinguish between **direct objects, indirect objects** and **object complement clauses**, for instance, *she hit it, she gave him the ball*, and *she said that it is sad*, respectively. As, for instance, Hawkins [12] shows, these three types of objects exhibit increasing degrees of complexity.
Relative clauses, such as *the man who walks on the other side of the street is my uncle*, have been found to be good indicators of

suspected partner competence and linguistic proficiency; thus, in human-robot interaction speakers only use **relative clauses** if they are certain to be understood or if their partner uses them as well (Fischer [5]). We therefore take uses of relative clauses as an indicator for complexity here.

Embedding is a composite feature, combining all structures that can be embedded in the main sentence structure, such as relative clauses, object complement clauses, dependent main clauses, subclauses, appositions, infinitival complements, and subject clauses. In particular, we use the following definitions: Subclauses are subordinate clauses like *whenever he goes to school, he feels sick*. Appositions are added elements, such as *see the button, the red one*. An example for an infinitival complement is *she wants to go* and for a subject clause *what she really wants is love*.

Interpersonal function

A third property concerns the amount of social information used and the degree with which speakers involve their communication partner. One such feature concerns the sentence type, in particular, imperative, declarative, interrogative or infinitive mood. The declarative is generally used to make assertions. Furthermore, instructions by means of **declarative** sentences are very common, thus avoiding that the speaker directly imposes his or her wishes onto the communication partner, as it is the case with a simple **imperative**, such as, for instance, *move*! In German, imperatives are however often toned down by means of **modal particles**, sentence medial particles that serve politeness and grounding functions (cf. Fischer [6]). In the current data sets, the down-toned imperative occurs frequently in expressions with attention getting functions, such as *guck mal* (*look*).

In situations without a concrete addressee (cf. Halliday & Matthiessen [11]), or with a highly unfamiliar addressee, such as a computer or robot (cf. Fischer [5]), instructions and explications using the **infinitive** are very common, for instance: *den blauen nehmen*; this corresponds roughly to the English use of the gerund, as in, for example, *no smoking*.

Moreover, speakers can ask **questions** to involve their addressees, or they can use understanding **checks**, such as tag questions like *doesn't it* or *don't you* in English and *ne?* in German.

Also, personal pronouns are useful indicators of the relationship between speakers in a communicative situation. For instance, speakers may avoid addressing the partner, using the impersonal form *man* (*one*). Alternatively, speakers can address their partner using *du* (*you*), or they can refer to themselves with or without including the partner, using either *ich* (*I*) or *wir* (*we*). Similarly revealing regarding the degree with which the communication partner is involved is the use of the **vocative**, for instance, the partner's first name.

The absolute occurrences of these features, besides the verbosity features, were counted per person in these conditions and divided by the number of utterances used by this person. The numbers underlying the statistical comparison are thus the numbers occurring per number of speakers' utterances.

5. RESULTS

The analyses reveal that many linguistic features remain the same in all conditions. Especially regarding the amount of speaking, i.e.

the verbosity measures, there are no differences, and also most complexity measures yield very similar results across conditions. There are, however, significant differences between all three conditions, as well as differences that only hold between two of the three conditions, pointing to different effects of physical embodiment on the one hand and degrees of freedom on the other.

Table 1: ANOVA Results for all Linguistic Features; * = p< .05, ** = p< .001; ^{t} = p< .20

	Akachan	iCub I	iCub II	F
diversity	7.9 *(0.2)*	8.9 *(0.1)*	6.7 *(0.1)*	1.30383
MLU	8.4 *(2.7)*	8.2 *(4.0)*	6.0 *(2.3)*	1.740385 ^{t}
concrete	0.75 *(0.34)*	1.0 *(0.6)*	0.9 *(0.4)*	0.27753
abstract	0.05 *(0.06)*	0.06 *(0.09)*	0.02 *(0.02)*	0.80275
passive	0.41 *(0.08)*	0.42 *(0.04)*	0.008 *(0.01)*	0.61220
direct obj	0.03 *(0.03)*	0.46 *(0.2)*	0.25 *(0.1)*	49.97369**
ind.object	0.03 *(0.05)*	0.03 *(0.07)*	0.07 *(0.02)*	0.39219
embedding	0.19 *(0.09)*	0.22 *(0.17)*	0.11 *(0.10)*	1.42739
subclauses	0.11 *(0.07)*	0.13 *(0.10)*	0.07 *(0.05)*	1.18885
rel.clause	0.03 *(0.03)*	0.03 *(0.02)*	0.005 *(0.01)*	1.34224
copula	0.04 *(0.04)*	0.19 *(0.16)*	0.19 *(0.1)*	17.04208**
declarative	0.92 *(0.3)*	1.11 *(0.6)*	0.76 *(0.3)*	1.93717 ^{t}
infinitive	0.06 *(0.05)*	0.08 *(0.1)*	0.09 *(0.09)*	0.61691
imperative	0.02 *(0.03)*	0.17 *(0.07)*	0.08 *(0.07)*	2.24545 ^{t}
mod.particle	0.10 *(0.07)*	0.21 *(0.15)*	0.19 *(0.12)*	4.71764*
question	0.04 *(0.05)*	0.03 *(0.05)*	0.01 *(0.02)*	0.91876
check	0.001 *(0.0)*	0.002 *(0.0)*	0.02 *(0.04)*	3.29369*
vocative	0.01 *(0.03)*	0.04 *(0.05)*	0.16 *(0.14)*	16.84286**
du (*you*)	0.03 *(0.09)*	0.06 *(0.07)*	0.02 *(0.04)*	0.51796
man	0.20 *(0.2)*	0.10 *(0.2)*	0.15 *(0.2)*	1.13165
ich (*I*)	0.15 *(0.15)*	0.17 *(0.17)*	0.12 *(0.08)*	1.12149
wir (*we*)	0.04 *(0.06)*	0.12 *(0.15)*	0.04 *(0.07)*	2.87466

It turns out that the vocative, i.e. the use of the robot's name, differs significantly in all conditions (see Figure 4). The mean in Condition 1 is M= 0.01 (sd= 0.03), in Condition 2 M= 0.04 (sd=0.05) and in Condition 3 M=0.16 (sd=0.13), t (Condition 1 vs. 2) = -2.29, p < .03; t (Condition 1 vs. 3) = -5.69, p < .001; t (Condition 2 vs. 3) = -2.26, p < .05. The results show that the number of instances of the robot's name increases with increasing physical embodiment of the robot and with increasing degrees of freedom.

We turn now to paired comparisons between conditions. Figures 4 to 6 illustrate the different possible distributions between the linguistic features investigated.

5.1 Simulated versus physical robot (eyes only)

The first comparison concerns the role of physical embodiment, i.e. Condition 1 (Akachan) versus Condition 2 (iCub I). Several features can be found that are influenced by the robot's physical embodiment; these are the number of instances of the personal pronoun 'we' (Condition 1 M=0.04, sd=0.064; Condition 2:

M=0.12, sd=0.15, t=-2.29, p< .03), the number of modal particles (Condition 1 M=0.10, sd=0.07; Condition 2 M=0.21, sd=0.15, t=-2.73, p< .01), the number of direct objects (Condition 1 M=0.03, sd=0.03; Condition 2 M=0.46, sd=0.24, t=-9.72, p< .001), and the number of utterances containing a copula, i.e. a form of 'to be' as the main verb of the sentence (Condition 1 M=0.04, sd=0.04; Condition 2 M=0.21, sd=0.16, t=-4.99, p< .001, see also Figure 5). Furthermore, there is a tendency (t=1.96, p= .058) for more imperatives in Condition 2 (M=0.05, sd=0.07) compared to Condition 1 (M=0.02, sd=0.02).
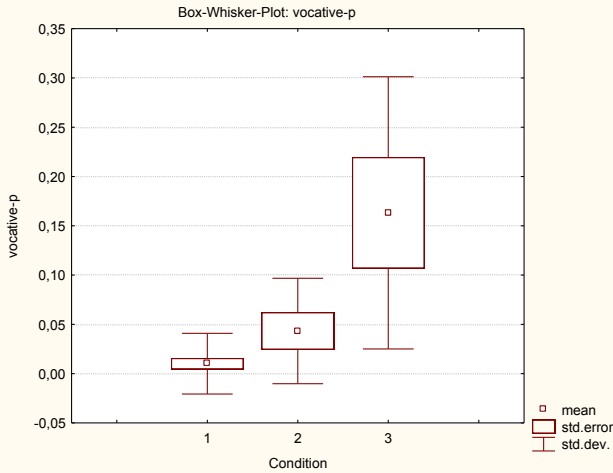

Figure 4: Results for the vocative in the three conditions

5.2 Physical robot (eyes only) versus physical robot (eyes and head)

The comparison between Conditions 2 and 3, i.e. between iCub I and iCub II, reveals the impact of the robot's head movements in Condition 3. Significant differences that point to an influence of the amount of degrees of freedom of the robot concern, besides the vocative, also the amount of passive constructions employed (Condition 2 M=0.04, sd=0.03; Condition 3: M=0.08, sd=0.01, t=2.19, p< .05). Thus, there are many more instances of the passive construction in Condition 2, the condition with iCub I, which is only using its eyes. Furthermore, there is a statistical tendency towards more direct objects (Condition 2 M=0.46, sd=0.24; Condition 3: M=0.26, sd=0.12, t=1.87, p< .09).
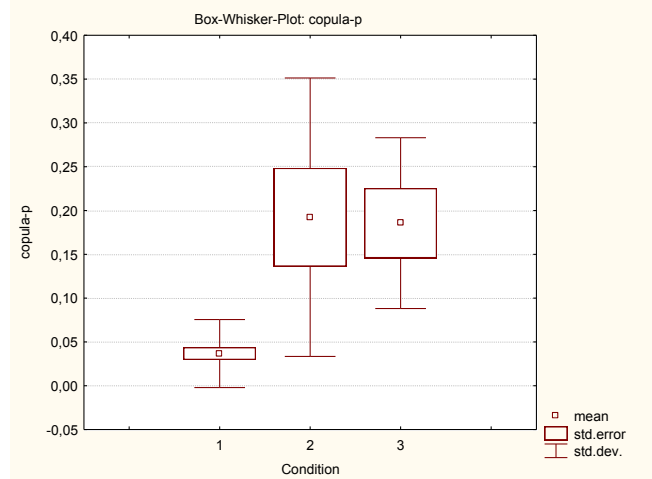

Figure 5: Results for the copula in the three conditions

5.3 Simulated versus physical robot (eyes and head)

Finally, we compare Condition 1, interactions with Akachan, with Condition 3, interactions with iCub II, who is moving both eyes and head. Significant differences concern, besides the vocative, also the amount of understanding checks employed (Condition 1 M=0.001, sd=0.004; Condition 3: M=0.017, sd=0.04, t=-2.35, p< .03). Furthermore, there are significantly more direct objects in Condition 3 than in Condition 1 (Condition 1 M=0.03, sd=0.03; Condition 3: M=0.25, sd=0.12, t=-9.35, p< .001). There is moreover a tendency for increased uses of the imperative in Condition 3 (Condition 1 M=0.016, sd=0.03; Condition 3: M=0.045, sd=0.07, t=-1.80, p= .08). There are also significantly more uses of the copula in Condition 3 than in Condition 1 (Condition 1 M=0.037, sd=0.04; Condition 3: M=0.19, sd=0.05, t=-6.43, p< .001). In addition, there are significantly more modal particles in Condition 3 than in Condition 1 (Condition 1 M=0.10, sd=0.07; Condition 3: M=0.19, sd=0.12, t=-2.44, p< .02). Finally, the evaluation also shows that users' mean length of utterance is significantly shorter in the third condition than in the other two conditions (Condition 1: M= 8.4, sd 2.7; Condition 2: M= 8.2, sd= 4.0; Condition 3: M= 6.0, sd= 2.3; t= 0.167, p< .05; see also Figure 6).

6. DISCUSSION

The distribution of the results is readily interpreted in the light of the functions the respective linguistic features fulfill in interaction. Notable is first that with the exception of the number of direct objects and the MLU, none of the linguistic features concerning linguistic complexity were found to differ significantly between the conditions. Instead, the differences concern interactional features of language: the use of the robot's name to direct its attention, the use of the inclusive personal pronoun 'we', the use of modal particles which indicate common ground between speaker and hearer (Diewald [5], Fischer [7]), the use of understanding checks and the tendency to use imperatives, a grammatical form that encodes the addressee directly. These features all concern interpersonal relationships, and the different amounts with which they are being produced indicate increasing amounts of interactivity over the three conditions, with the more

embodied robot iCub II being taken as a more serious communication partner than the simulated robot.
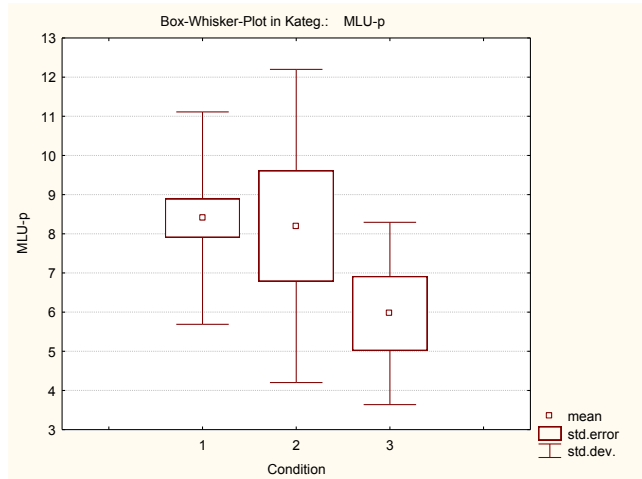

Figure 6: Results for Mean Length of Utterance (MLU) in the three conditions

The use of the copula and the direct object do not seem to fit in this pattern; however, in the current dialogs, the copula is used to introduce an object before something is asserted about it. For instance, participants say something like "iCub, this is a lamp, and to switch it on, you pull the string." The use of the copula 'is' is thus an indicator for the use of a tutoring strategy by means of which a task is decomposed into smaller steps. The same holds for the direct objects: Sentences with a single agent and a direct object are very straightforward and easy to understand and they constitute the most common utterances containing a verb in child-directed speech (Zeschel [32]).

The passive however is often an indicator for problems concerning the interpersonal relationship because it focuses on an event without stating the agent, for instance: "a lamp is switched on by pulling the string." In this utterance using the passive, a direct address of the communication partner is avoided (in contrast to, for instance, "you pull the string" or even "pull the string"). Increased use of the passive thus indicates that participants are uncomfortable with their communication partner.

The functional analysis of the choices made by the participants thus points to differences in the interpersonal relationships participants build up with the robots. This finding corresponds to previous findings concerning the 'enjoyability' or 'pleasantness' of physically embodied robots, which have been found in post-experimental questionnaires. However, while post-experimental questionnaire studies do not allow any conclusions with respect to why participants find the interactions more enjoyable, the current results show that these differences are likely to be due to the interpersonal relationship with the communication partner such that the robot is accepted as of the same kind ('we'), that tutor and robot share common ground (modal particles) and that the robot is an independent entity with a name who can be summoned in order to draw its attention (vocatives).

Furthermore, the results show that not only physical embodiment matters but also the robot's degrees of freedom; the linguistic features differing between the simulation and the iCub with head movement condition show that a physically embodied robot that uses its body at least to some degree is considered to be more likely to profit from the tutoring than a simulated robot or a robot that does not make use of its body in a way that suggests task fulfillment. Thus, linguistic differences between the iCub with head condition and the other two conditions concern, besides interpersonal differences, also task decomposition and presentation (mean length of utterance, direct objects, understanding checks). Thus, while the degrees of freedom of the agent in the study by Lusk & Atkinson [23] did not contribute to task completion and hence had no impact on the interaction, whether the robot can move its body is crucial concerning the relevance of teaching it to stack cups onto each other. Thus, the suggestion made here is that the number of degrees of freedom matters to participants in this study because the robot's movements are indicative of the robot's capabilities and thus its credibility as a partner for the task at hand.

7. CONCLUSION
In this study, we have employed a new methodology to address the effect of different degrees and types of embodiment of a given system on human-robot interaction. The method employed allowed us to identify significantly different behaviors in interaction with the three systems investigated. Moreover, the different linguistic behaviors observed could be related to different ways of understanding the robotic agents and thus to the factors determining the influence of degrees of a robot's embodiment on human-robot interaction. In particular, it was found that the two factors studied influence both the interpersonal relationship between human user and the robot and the amount of tutoring the robot received.

8. DESIGN IMPLICATIONS
The current investigation has shown that not only the robot's physical embodiment, but also its degrees of freedom influence human-robot interaction. The robot behaviors displayed were found to be relevant especially in the light of the tasks it is meant to fulfill. Thus, the robot should use its degrees of freedom in a way that is in accordance with its capabilities.

ACKNOWLEDGMENTS

REFERENCES
[1]   Bainbridge, W.A., Hart, J.W., Kim, E.S. and Scasselati, B. 2010. The Benefit of Interactions with Physically Present Robots over Video-Displayed Agents. *International Journal of Social Robotics* 1-2.

[2]   Cross, T., Johnson-Morris, J.E. and T.G. Nienhuys. 1980. Linguistic Feedback and Maternal Speech: Comparisons of Mothers Addressing Hearing and Hearing-impaired Children. *First Language* 1, 3:163-189.

[3]   Dautenhahn, K., B. Ogden and T. Quick. 2002. From Embodied to Socially Embedded Agents – Implications for Interaction-aware Robots. *Cognitive Systems Research* 3: 397-428.

[4]   Diewald, G. 2006. Discourse Particles and Modal Particles as Grammatical Elements. In Fischer, K. (ed.): *Approaches to Discourse Particles*. Amsterdam: Elsevier.

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

[6]   Fischer, K. 2007. Grounding and common ground: Modal particles and their translation equivalents. In A. Fetzer and K. Fischer (eds.), *Lexical Markers of Common Grounds*. Studies in Pragmatics 3. Amsterdam: Elsevier.

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

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

[9]   Foth, K., Menzel, W. and Schröder, I. 2000. A Transformation-based Parsing Technique with Anytime Properties. *4th Int. Workshop on Parsing Technologies,* IWPT-2000, pp. 89-100.

[10]   Fussell, S.R., Kiesler, S., Setlock, L.D. and Yew, V. 2008. How people anthropomorphize robots. *Proceedings of Human-Robot Interaction 2008* (pp. 145-152). NY: ACM Press.

[11]   Halliday, M.A.K. and Matthiessen, C. 2004. *Introduction to Systemic Functional Grammar*. London: Arnold.

[12]   Hawkins, J.A. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press.

[13]   Hoffmann, L. and Krämer, N.C. 2011. How Should an Artificial Entity be Embodied? Comparing the Effects of a Physically Present Robot and its Virtual Representation. *Proceedings of the HRI 2011 Workshop on Social Robotic Telepresence*, Lausanne, Switzerland, pp. 14-20.

[14]   Itti, L., Koch, C. Niebur, L. et al. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259.

[15]   Kidd, C. and Breazeal, C. 2004. Effect of a Robot on Uer Perceptions. *IEEE/RSJ International Conference on Intelligent Robots nd Systems*, pp. 3559-3564.

[16]   Kiesler, S., Powers, A., Fussell, S. and Torrey, C. 2008. Anthropomorphic Interactions with a Robot and Robot-like Agent. *Social Cognition* 26, 2: 169-181.

[17]   Koda, T. and P. Maes. 1996. Agents with Faces: The Effect of Personification. *Fifth IEEE International Workshop on Robot and Human Communication*.

[18]   Komatsu, T. and Abe, Y. 2008. Comparing an On-Screen Agent with a Robotic Agent in Non-Face-to-Face Interactions. In Prendinger, H., Lester, J. and Ishizuka, M. (eds.), *IVA 2008*. Berlin, Heidelberg: Springer, pp. 498-504.

[19]   Komatsu, T. and Kuki, N. 2009a. Can Users React toward an On-Screen Agent as if they are Reacting toward a Robotic Agent? *HRI'09*, La Jolla, CA, pp. 217-218.

[20]   Komatsu, T. and Kuki, N. 2009b. Investigating the Contributing Factors to Make Users React toward an On-Screen Agent as if they are Reacting toward a Robotic Agent. *18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan.

[21]   Lohan, K., Gieselmann, S., Vollmer, A.L., Rohlfing, K. and B. Wrede. 2010. Does Embodiment Affect Tutoring Behavior? *Proceedings of ICDL'10.*

[22]   Lopes, M., Bernardino, A., Santos-Victor, K. Rosander, and C. von Hofsten,. 2009. Biomimetic Eye-Neck Coordination. *ICDL 2009*.

[23]   Lusk, M.M. and R.K. Atkinson. 2007. Animated Pedagogical Agents: Does their Degree of Embodiment Impact Learning from Static or Animated Worked Examples? *Applied Cognitive Psychology* 21: 747-764.

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

[25]   Nagai, Y. and Rohlfing, K.J. 2009. Computational analysis of Motionese toward scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development* 1: 44-54.

[26]   Nass, C. 2004. Etiquette Equality: Exhibitions and Expectations of Computer Politeness. *Communications of the ACM* 47, 4: 35-37.

[27]   Roy, B., Frank, M.C. and Roy, D. 2009. Exploring word learning in a high-density longitudinal corpus. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Amsterdam, Netherlands.

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

[29]   Sproull, L., Subramani, R., Kiesler, S., Walker, J., and Waters, K. 1996. When the interface is a face. *Human-Computer Interaction* 11, 97-124.

[30]   Takayama, L. 2009. Making Sense of Agentic Objects and Teleoperation: In-the-moment and Reflective Perspectives. *Late Breaking Results of Human Robot Interaction* (HRI), San Diego, CA, pp.239–240.

[31]   Wainer, J., Feil-Seigfer, D.J., Shell, D.A. and Mataric, M.J. 2007. Embodiment and Human-Robot Interaction: A Task-Based Perspective. *16th IEEE International Conference on Robot and Human Interactive Communication*, Jeju, Korea.

[32]   Zeschel, A. forthcoming. Basic Argument Structure Constructions in Child-Directed Speech. *Journal of Child Language.*

[33]   Ziemke, T. 2003. What's that Thing Called Embodiment? In: Alterman and Kirsh (eds.) *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 1134-1139). Mahwah, NJ: Lawrence Erlbaum.