

Does the Appearance of a Robot Affect Users' Ways of Giving Commands and Feedback?

Anja Austermann, Seiji Yamada, Kotaro Funakoshi and Mikio Nakano

Abstract—Our study compares users' interaction with a humanoid robot and a dog-shaped pet-robot. We conducted a user study in which the participants had to teach object names as well as simple commands to either the humanoid or the pet-robot and give feedback to the robot for correct and incorrect performance. While we found, that the way of uttering commands rather depends on personal preference than on the robots' appearance, the way of giving positive and negative feedback differed significantly between both robots: We found that for the pet-robot users gave reward in a similar way as giving reward to a real dog by touching it and commenting on its performance by uttering feedback like "well done" or "that was right". For the humanoid, users typically did not use touch as a reward and rather used personal expressions like "thank you" to praise the robot. Our findings suggest that users actually rely to some degree on the appearance of a robot as a cue for deciding how to interact with it.

I. INTRODUCTION

When humans interact with other humans or with their pets they tend to adapt their way of speaking and interacting to their interaction partner. For example, people talk to adults in a more elaborated way than to small children, and they pet their dog as a reward while they would rather say "thank you" when their colleague has done them a favor. Moreover, they speak more slowly and clearly, when they assume their communication partner is not understanding them well.

We assume that similar mechanisms also affect how people interact with robots. Especially the appearance of a robot and its resemblance to familiar creatures or objects can be an important factor which helps a human to anticipate the capabilities of a robot and decide how to interact with it. The results from our research can help inform the design choices that roboticists make when considering what type of interaction they want with their robots.

II. RELATED WORK

In recent years, there have been various studies [8] [4] [5] [3] investigating the effect of a robot's appearance on the interaction with a user. However, most studies concerning the appearance of robots rather deal with the uncanny valley effect [2] and users' impression of robots than with the effect of a robot's appearance on its user's communicative behavior.

Kanda et al. [4] conducted a study with two different humanoid robots and showed that different appearances of

the robots did not affect the participants' verbal behavior but did affect their non-verbal behavior such as distance and delay of response. They explain the observed differences by impressions, such as novelty, safety, familiarity and activity as well as attributions, such as whether the robot is respected as a conversation partner.

Kriz et al. [8] investigated users' conceptualizations of robots by analyzing the way the users talked to the robot. They compared features of robot-directed speech to how humans talk to infants or adult non-native speakers. They found that the participants spoke more loudly, raised their pitch, and hyperarticulated when they spoke to the robot. This behavior is typical when the conversation partner is assumed to have low linguistic competence. However, they did not speak in easier sentences, which suggests, that they believed that the robot has almost humanlike cognitive capabilities.

Goetz et al. [5] investigated users' attribution of capabilities depending on the appearance of a robot. They created images of more or less human-like looking robots and had participants judge their suitability for different tasks. They found that people systematically preferred robots for jobs when the robot's human-likeness matched the sociability required in those jobs. They also found in a second user study with a humanoid robot, that playful or serious demeanor of the robot affects the compliance of the participants. The participants performed a playful task longer, when the instructing robot showed a playful demeanor while the participants performed a serious task longer, when the robot behaved more seriously.

Similar results were obtained by Hegel et al. [3] who found that the appearance of robots affected users' attribution of possible applications. They conducted a user study in which the participants were asked to match videos of twelve robots to thirteen different categories of applications. Especially the perceived human-likeness or animal-likeness affected which tasks the participants considered suitable for each robot. While the participants considered human-like robots for fields like healthcare, personal assistance, security and business, they considered animal-like robots as companions, entertainers, toys, and robotic pets.

III. OUTLINE OF THE STUDY

We conducted a user study on how participants give commands and feedback to a pet-robot and a humanoid. As a pet-robot, we used a dog-shaped robot, which has roughly the size of a cat or a small dog. The humanoid robot is 1,20 m tall, which is about the size of a 8-year-old child. Both

A. Austermann is with the Graduate University for Advanced Studies, Tokyo, Japan anja@nii.ac.jp

Seiji Yamada is with the Graduate University for Advanced Studies and the National Institute of Informatics, Tokyo, Japan seiji@nii.ac.jp

Kotaro Funakoshi and Mikio Nakano are with the Honda Research Institute Co., Ltd., Wako, Japan funakoshi@jp.honda-ri.co.jp, nakano@jp.honda-ri.co.jp

TABLE I
COMMANDS THAT WERE USED IN THE TRAINING TASK.

Command	Parameters	Example sentence
move	object, place	Put the ball into the box.
bring	object	Bring me a coffee, please.
clean	object	Please clean up the carpet.
switch on	object	Robot, switch on the light.
switch off	object	Switch off the radio.
call	object	Please make a phone call to Rita.
charge battery	-	Recharge your battery.
show status	-	What is your status?

robots are shown in Fig. 2. The goal of our study was to find differences and similarities in user behavior when the participants give commands and feedback to the pet-robot and the humanoid. The user study, described in this paper, is a part of our work on learning commands and feedback for human-robot interaction [1].

Each participant interacted with either the humanoid or the pet-robot and instructed the robot to perform different typical household tasks like bringing a coffee, switching on the light or the TV, tidying up etc. and gave feedback to the robot for correct or incorrect performance.

A. The “Virtual Living Room”

In order to avoid time-consuming and error-prone task execution in the real world and because of the different physical capabilities of the two different robots, we implemented a “virtual living room”. The tasks as well as the actions of the robot were visualized on a large screen. The robot was placed in front of the screen and used motion and speech to inform the user which action it is currently performing in the virtual living room. The robots’ actions and pointing direction were also visualized in the living room scene with a hand or paw icon and the scene changed in response to the actions of the robot. Based on these cues the participants could easily understand the relation between the robot’s motions and the changes happening in the scene. While the robots differed in shape and size we kept all other parameters as similar as possible, using the same synthesized speech utterances, similar gestures, same simulated learning rate, almost same position of the robot in relation to the user etc.

During the training, the user could figure out by looking at the scene, what command to give to the robot next. We used a graphical representation of the scene without any text, in order to avoid influencing the participants’ wording when giving commands to the robot.

Table I shows the list of all commands, that were used in the task and sample utterances for each command. The users were not instructed in advance, which commands they had to teach to the robot but were asked to infer which commands were appropriate by looking at the virtual living room scene.

B. The Training Phases

One experiment with one participant comprised two successive training phases. In the first phase, the user had to

teach the names of eighteen different objects to the robot. The robot pointed at objects on the screen and a spotlight as well as a pointing arrow was shown in the living room scene to make it easier for the user to understand the robot’s pointing direction. The robot then asked “What is that” (“kore ha nan desu ka?”) to prompt the user for an object name. After the user had uttered an object name, the robot continued with the next object.

We asked the users to only utter the object names without any additional words. This was a requirement for our learning algorithm. Because of this restriction, the speech, recorded in the first training phase was not evaluated.

In the second phase, different scenes were shown on the screen for learning commands. As described above, each scene visualized a task, that had to be performed by the robot, in a way that the user could understand which instruction would be suitable. The robot looked at the user while it was waiting for an instruction. After the user had uttered an instruction, the robot either performed correctly or incorrectly. For example when the user uttered a command like “Can you bring me a coffee”, the robot would perform correctly by pointing at the screen and making an appropriate gesture. As a visualization, the robot hand icon on the screen would move to the table to put the coffee cup there and then the robot would say “Here you are” (“douzo”). For an incorrect performance, the robot would, for example switch off the light instead of bringing a coffee.

After executing the user’s command either correctly or incorrectly, the robot looked at the user to wait for feedback. After receiving positive or negative feedback, the robot confirmed by either thanking for positive feedback or confirming that it understood the negative feedback. As the robot could not actually understand the given feedback during the training task, it used the heuristics, that feedback, given after a correct performance, is positive feedback and feedback, given after an incorrect performance is negative feedback. This expectation almost always agreed with the actual feedback given by the user. After the robot had acknowledged the user’s feedback, the next scene was shown on the screen, so that the user could continue teaching the next command.

As the robot had direct access to the task server and could request scenes for a certain command from the task server, it was able to make the user give commands with a certain meaning and also provoke positive or negative feedback by correct or incorrect performance. Therefore it did not need to apply speech recognition in order to understand the user’s actual utterances and could run autonomously without remote-control.

The robot was equipped with a voice activity detection, so that it was able to react, when the user uttered a command or feedback. Any speech utterance or touch, occurring after a new scene had been shown, was assumed to be a command. Every utterance or touch, that occurred after the robot had executed a command, was expected to be feedback.

When executing any of the commands, the robot performed a specific gesture. The gestures were selected so that

we could implement them in a similar way on the four-legged pet-robot and the humanoid.

Sample virtual living room scenes for prompting the user to give a command to the robot are shown in Fig. 1. In the first scene, the robot asks the user to name the “audioplayer” object. In the second scene, the user is expected to tell the robot to switch the light on. In the third scene, the user is expected to make the robot switch off the television. The white texts in the images show our internal representations of commands and were not shown to the user during the training task. Details on the implementation of the system and the use of the training tasks for actually learning to understand commands and feedback are given in [1].

IV. ASSUMPTIONS

Based on the schema theory [7] in psychology, which suggests that people use schemata of familiar objects and situations to understand and handle unfamiliar situations, we assumed that users are likely to interact with a pet-robot in a similar way as with a real dog, while interaction with a humanoid was expected to resemble more to the interaction with a human.

Moreover, we assumed that the participants were likely to conclude that the humanoid is more intelligent than the pet-robot, based on its humanlike appearance. This might lead to higher expectations and to adaptations such as a more elaborated speaking style, more politeness, more explanations etc. when interacting with the humanoid. Details on our expectations as well as the actually observed interaction are given in the results section.

V. EXPERIMENTAL SETTING

We have conducted a user study with 16 participants aged from 22 to 52. Ten participants (7 males, 3 females) interacted with the humanoid and six participants (4 males, 2 females) interacted with the pet-robot for roughly 45 minutes. The language, used in the experiments, was Japanese. All participants were employees of the Honda Research Institute Japan. 8 participants keep a pet or have kept a pet and 5 participants have experience in keeping dogs.

Fig. 2 shows the experimental setting. The participants were asked to sit at a table in order to avoid excessive changes of position during the experiment. This was necessary because we also recorded video data for analysis and for gesture recognition. The robot was placed to the right of the participant, close enough that all participants could easily reach it with the hand to touch it. As the pet-robot was a lot smaller than the humanoid, it was placed on the table, so that the participants could reach it easily. The participants were equipped with a headset microphone to record audio data. Video data was recorded using a stereo camera which was placed above the screen.

The participants were given explanations about the two training phases. In the first phase, they were asked to name the objects that the robot was pointing at. In the second phase, they were instructed to give commands to the robot and to give positive feedback if the robot reacted correctly

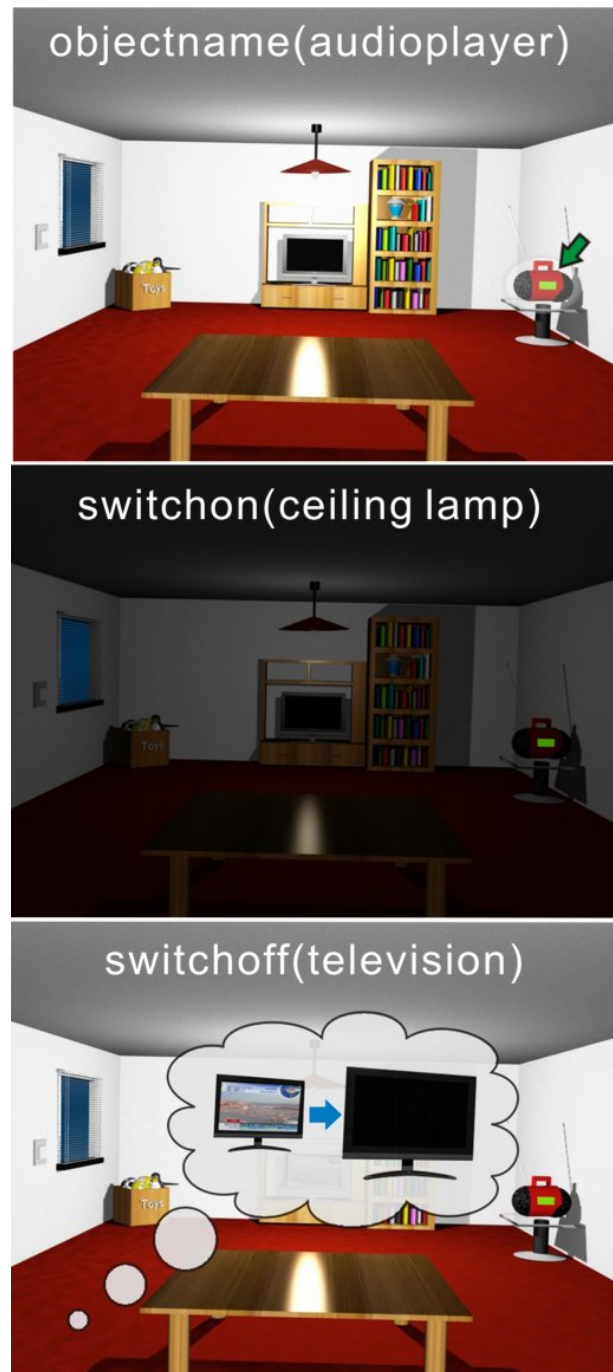


Fig. 1. Sample Scenes from the “Virtual Living Room”

and negative feedback if the robot reacted incorrectly. They were instructed to give commands and feedback in any way they liked by speech, gesture and touch. The participants had to teach each object name and each command ten times. As the duration of an experiment was relatively long and the users were required to talk a lot, there was five minute break between the training sessions for object names and commands.



Fig. 2. Experimental Setting.

TABLE II
USERS' EVALUATION OF THE TRAINING TASK

Question (5: fully agree - 1: do not agree)	Humanoid	Pet-robot
I enjoyed teaching the robot through the given task	3.5 (0.8)	4 (0.8)
The robot understood my feedback	3.6 (0.9)	4.3 (1.1)
The robot learned through my feedback	3.2 (1.3)	4.3 (0.5)
The robot adapted to my way of teaching	3.2 (1.1)	3.8 (1.3)
I was able to instruct the robot in a natural way	3.6 (1.1)	3.5 (1.5)
The robot took too much time to learn	3.6 (1.4)	2.7 (0.9)
The robot is intelligent	2.7 (1.3)	2.8 (1.5)
The robot behaves autonomously	2.7 (1.4)	2.8 (0.9)
The robot behaves cooperatively	3.7 (0.8)	3.3 (0.7)

VI. RESULTS

In our user study, we obtained two different kinds of results: We asked the participants to answer a questionnaire about their subjective impression of the interaction and we annotated the data, which was recorded during the interaction to find objective similarities and differences in the participants' behavior. We used the T-test to determine the statistical significance of the observed differences.

A. Questionnaire results

From the results of the questionnaire, which are shown in table II we can see a slight tendency towards more positive ratings for the interaction with the pet-robot. However, none of the differences is statistically significant.

B. User behavior

We analyzed different aspects of the participants' commands and feedback that we assumed to be related to the perceived intelligence and human-likeness of the robot. We compared the speaking speed (in seconds per word) and the number of words per command/feedback, as we assumed that people talk slower and in simpler sentences, when they

TABLE III
TYPES OF COMMANDS USED IN THE INTERACTION WITH THE HUMANOID AND THE PET-ROBOT

Type	Humanoid	Pet-Robot
Plain commands	75.01 (14.00)	60.83 (41.04)
Polite commands	9.86 (10.88)	26.23 (41.99)
Questions in commands	10.23 (3.51)	8.34 (6.73)
Implicit commands	3.40 (4.82)	4.10 (7.23)
Parameters left out	6.78 (2.25)	4.13 (4.77)
Explanations in commands	1.81 (3.90)	0.95 (2.32)

All values in percent, value in brackets is the standard deviation

consider the robot less intelligent. However, we found, that the length of commands was almost the same for both robots. An average command for the humanoid was 3.75 (sd=0.42) words long, while an average command for the pet-robot was 3.72 (sd=0.71) words long. The speaking speed was also similar for the pet-robot with 0.45 (sd=0.09) seconds per word, and the humanoid with 0.42 (sd=0.07) seconds per word. This is in line with the participants' subjective evaluation of the robots' intelligence, shown in Table II.

C. Multimodality

During the interaction with both robots, we did not observe pointing gestures from any of the users. A possible explanation is that all objects were very easy to distinguish verbally, so that pointing gestures would have been redundant. We observed touch-based rewards for only one out of ten participants for the humanoid but for five out of the six participants who interacted with the pet-robot. As touch is frequently used with real dogs, we assume that users considered touch to be appropriate for giving feedback to a pet-robot because of its dog-like appearance.

D. Verbal Commands

We analyzed how many commands had explanations or polite expressions and how many commands were phrased as a question. We estimated that users might be more polite, explain more and use more questions when talking to a humanoid robot, while they rather give plain commands to a dog-like robot. We considered commands that contain words like "kudasai", "kureru?", "moraeru?" etc., which are similar to the English word "please" as polite commands. We also analyzed, how many commands were implicit ones like saying "it is too dark here" to make the robot switch the light on, and in how many commands some expected parameters were left out like in "put away the toy car" instead of "put the toy car into the box", because we assumed that this kind of verbal behavior might be related to the perceived intelligence of the robot.

The results can be found in Table III. The values do not add up to 100% because not all types of commands are mutually exclusive (e.g. a polite command can have parameters left out).

TABLE IV
TYPES OF FEEDBACK USED IN THE INTERACTION WITH THE HUMANOID
AND THE PET-ROBOT

Type	Humanoid	Pet-robot
Personal	52.78 (17.99)	24.83 (27.41)
Performance evaluation	38.39 (18.28)	70.02 (28.16)
Explanations	11.10 (14.29)	3.56 (3.90)

All values in percent, value in brackets is the standard deviation

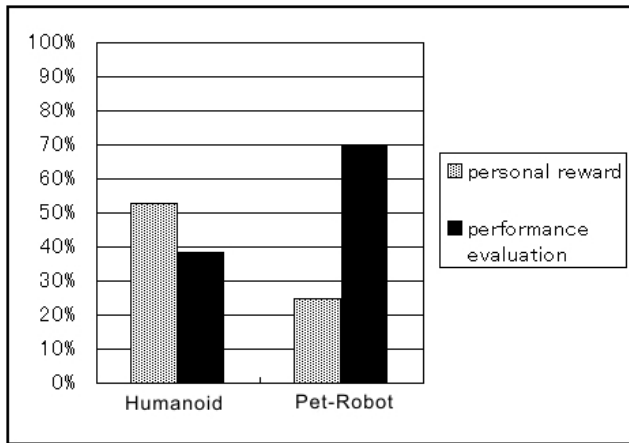


Fig. 3. Difference in Feedback for the Humanoid and the Pet-robot.

While we observed quite different utterances for different users, the differences seemed to be rather caused by personal preferences, than by the appearance of the robots. This assumption is supported by the high standard deviations between participants. None of the observed differences was statistically significant.

E. Verbal Positive and Negative Feedback

We distinguished three different types of feedback: Personal rewards like “Thank you”, which emphasize, that the robot has done something for the user, feedback which directly comments on the performance of the robot, like “Well done.” or “That was wrong.” and explanations used as rewards like “That is not a toy car, it is a ball.” or “That is a toy car.”. The usage of different rewards for the humanoid and the pet-robot is shown in table IV. We found statistically significant differences for the usage of personal rewards ($df=14$, $t=2.48$, $p=0.026$) and rewards, which comment on the robots’ performance ($df=14$, $t=2.75$, $p=0.016$). While the participants usually gave feedback like “well done (yoku dekimashita)” or “good (ii yo)” to the pet-robot, they used more personal rewards like “Thank you (arigatou)” for the humanoid, especially for positive reward. Fig. 3 shows the differences in user feedback given to the humanoid and the pet-robot. While the participants gave more explanations when talking to the humanoid, especially for negative rewards, the difference between both robots was not significant.

F. Behavior Changes over Time

We also investigated the changes in user behavior over time by comparing the commands and feedback, the participants gave in the first five minutes of the command learning phase to the commands to the feedback given throughout the whole experiment and to the last five minutes of the experiment.

We did not find any significant changes in commands given to both robots over time. We also did not observe significant changes in the feedback given to the pet-robot. However, we could observe two marginally significant changes in the feedback given to the humanoid: The amount of explanations for negative feedback was marginally significantly lower ($p=0.071$, $t=2.06$, $df=9$) at the beginning of the experiments than it was throughout the whole experiment. While at the beginning of the experiment only 26.98% ($sd=32.32\%$) of the negative feedback for the humanoid contained an explanation, it was an average of 34.57% ($sd=35.87\%$) during the whole experiment and went up to 75.00% ($sd=35.36\%$) at the end of the experiment.

We also observed a marginally significant increase ($p=0.091$, $t=1.90$, $df=9$) in personal feedback given to the humanoid comparing the first five minutes of the command learning to the whole command learning phase. Overall, the percentage of personal feedback increased from 34.85% ($sd=22.62\%$) in the first five minutes to 61.92% ($sd=24.60\%$) in the last five minutes, while the average was 52.78% ($sd=17.99\%$).

Similar trends toward more personal feedback and more explanations for negative feedback were also found for the pet-robot. However, the statistical significance of these trends could not be confirmed. Fig. 4 compares the feedback given during the whole task to the feedback given during the first five and last five minutes of the task. For the interaction with the pet-robot we did not observe any explanations within the last five minutes of the training. This is because we looked at explanations accompanying negative feedback when the robot made mistakes. Due to the simulated learning, the robot made less mistakes towards the end of the training. As the amount of explanations was generally lower for the pet-robot, we could not observe any negative feedback with explanations in the last five minutes of the experiments with the pet-robot.

VII. DISCUSSION AND CONCLUSION

In our experiments, we observed less than expected differences in participants’ behavior toward the pet-robot and the humanoid. While especially the way of uttering commands seems to depend rather on the personal preferences of the user, than on the appearance of the robot, we found robot-dependent differences in the feedback, given by the participants. The most obvious one was the frequent use of touch for giving feedback to the pet-robot, while touch was almost not used for the humanoid. Moreover, we found, that users tended to give personal feedback like “Thank you” to the humanoid, while they rather commented on the performance for giving feedback to the pet-robot. These findings suggest

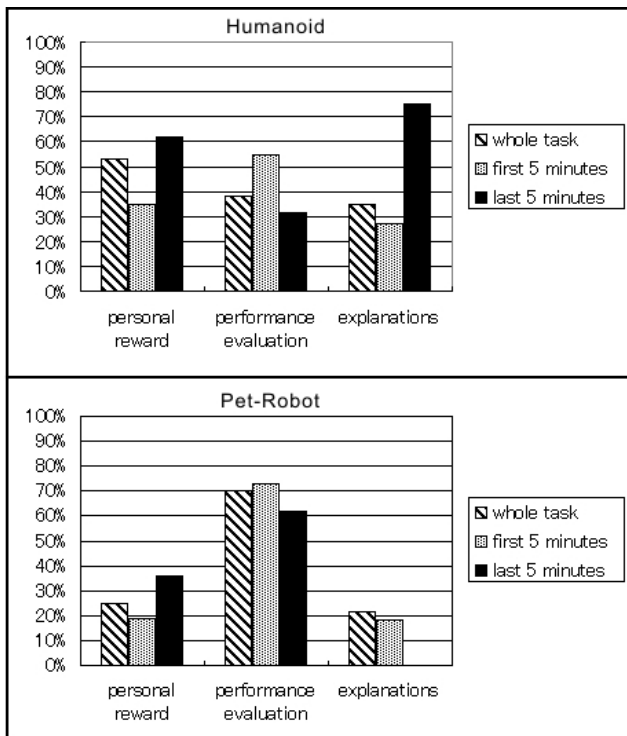


Fig. 4. User Feedback Changes over Time.

that people actually use their experience with real dogs as a guideline when giving feedback to the pet-robot.

When interacting longer with the humanoid, people started to give more explanations when the robot performed incorrectly and also gave more personal reward. While the results are only marginally significant and hard to interpret, one explanation may be, that the perception of the humanoid robot as an intelligent interaction partner increases when the robot shows learning capabilities and improves its performance during the experiment. Similar tendencies could be observed with the pet-robot. However, these tendencies were not statistically significant.

The users' subjective evaluation did not reveal significant differences between the humanoid and the pet-robot. As both robots were programmed to behave in the same way on the same task, we assume that the users' impression of the robots' behavior on the given task depends rather on their actual performance than on their appearance.

There are different possible explanations, why no significant differences were observed for giving commands. One of them is that both robots used speech to communicate with the user. As speech is a typical human modality of interacting, differences might have been stronger, if the pet-robot had communicated with the user in a more dog-like non-verbal way. As there was no significant difference in users' evaluation of both robots' intelligence, users may have considered similar types of commands acceptable for both robots.

While the initial T-test shows some interesting differences between the feedbacks that were statistically significant it

would be necessary to study more people including participants from outside Honda Research Institute to confirm this trend scientifically. Moreover, further experiments would be necessary to confirm whether the trend that we found in our experiments with one particular humanoid and one particular pet-robot and a special training task can actually be generalized to other types of humanoids or pet-like robots and to more general tasks. As discussed in section II, previous literature suggests, that depending on their appearance, user-behavior can vary for different types of humanoid robots and presumably the same is true for different pet-robots.

In our future work, we are planning to further analyze the variability and robot-dependence of given commands and feedback. The results will be applied to improve our method for learning to understand commands and feedback through a training task.

REFERENCES

- [1] A. Austermann, S. Yamada: "Learning to Understand Parameterized Commands through a Human-Robot Training Task", IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'09), (2009), 757-762
- [2] K. F. Mac Dorman: "Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it?" CogSci-2005 Workshop: Toward Social Mechanisms of Android Science, (2005), 106-118
- [3] F. Hegel, M. Lohse, B. Wrede: "Effects of Visual Appearance on the Attribution of Applications in Social Robotics", IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'09), (2009), 64-71
- [4] Takayuki Kanda, Takahiro Miyashita, Taku Osada, Yuji Haikawa, and Hiroshi Ishiguro: "Analysis of Humanoid Appearances in Human-Robot Interaction", IEEE TRANSACTIONS ON ROBOTICS, VOL. 24, NO. 3, (2008), 899 - 906
- [5] J. Goetz, S. Kiesler, and A. Powers, Matching robot appearance and behavior to tasks to improve human-robot cooperation, IEEE Workshop on Robot and Human Interactive Communication (ROMAN'03), (2003), 55 - 60
- [6] T. Komatsu, S. Yamada: "Effect of Agent Appearance on People's Interpretation of Agent's Attitude", CHI-2008 Work-in-Progress, (2008), 2919 -2924
- [7] W. F. Brewer: "Bartlett's Concept of the Schema and Its Impact on Theories of Knowledge Representation in Contemporary Cognitive Psychology." Bartlett, Culture and Cognition, ed. Akiko Saito. Hove, Eng.: Psychology Press, (2000), 69-89
- [8] Sarah Kriz, Gregory Anderson, J. Gregory Trafton: "Robot-Directed Speech: Using Language to Assess First-Time Users' Conceptualizations of a Robot", 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010), (2010), 267-275