Addressing User Experience and Societal Impact in a User Study with a Humanoid Robot

Astrid Weiss[1], Regina Bernhaupt[2], Manfred Tscheligi[1] and Eiichi Yoshida[3]

Abstract. Using a preliminary exploratory case study the presented work investigates the feasibility of methods for the evaluation of user experience factors in human-humanoid interaction as well as measurements of societal impact in user studies. The case study is based on two tasks participants had to perform with the HRP-2 robot. The robot was controlled by the participants via speech commands to pick up (task1) and put down (task2) an object onto another place. Main goal of the case study was to explore the methodological concept on how to measure novice users' experiences during the collaboration with the humanoid robot HRP-2 via speech commands and if the general attitude towards robotics changes because of the interaction with the robot. To address users' experiences during the user study, participants were asked after each task to state their thoughts and feelings they had during the interaction (retrospective think aloud). Furthermore, participants were interviewed by means of two validated standardized questionnaires (NARS and AttrakDiff) and an especially developed questionnaire. The preliminary results show that retrospective think aloud is a good way to gather qualitative data on users' experiences. Furthermore a final interview on societal impact of humanoid robots gave insights into how novice users imagine a future society with robots.

1 INTRODUCTION

"User Experience" (abbr. UX) is a concept in the area of Human-Computer Interaction (HCI) referring to the quality of the experience a person has when interacting with a specific design [6] and technology in general [10]. Research on UX currently has three facets [19]: First, UX is seen as a concept that goes beyond the instrumental, e.g. the standard measurements of usability like efficiency and effectiveness have to be extended with a concept like fun. Second, UX is related to emotion and affect referring to the positive aspects of an experience and how to support them (shifting from the perspective of trying to avoid negative consequences). Third, UX represents a concept that is related to a product and user experiences are embedded in a specific situation. Interaction goals, intra-psychological dispositions, the environment, involved people and the product itself have a significant impact on how users experience the product. Each experience has a discrete beginning and an ending. During this time span a user experiences emotions, which are heavily influenced by the introduced components.

1 HCI&Usability Unit, ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria. Email: {firstname.lastname}@sbg.ac.at.

2 IHCS-IRIT, University Paul Sabtier, 118 Route de Narbonne, 31062 Toulouse Cedex 9. Email: regina.bernhaupt@irit.fr

3 National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki 305-8564, Japan; Email: yoshida@aist.go.jp;

There is an increased interest in Human-Robot Interaction (HRI) to establish a positive experience for humans interacting with (humanoid) robots, especially in home settings [28]. Focusing on humans interacting with robots, a possible definition of the user's experience can be based on Alben's [17] general definition that UX includes "aspects of how people use an interactive product: the way it feels like in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it".

UX in HRI currently is evaluated using methodologies from the area of HCI. The applicability of these methods is still unclear as the validation of UX evaluation methods is difficult. Especially when interacting with a robot, individual user experiences might be heavily influenced by the individual's general attitude and the overall societal opinion.

Our general goal is to understand to what extent user experience is related to the overall acceptance of robots in society (societal impact). A first step is to investigate a methodological mix (as used in the area of HCI) to understand if HCI methods can be fruitfully used to understand user's experiences when interacting with a robot. The case study thus addressed the thoughts and feelings of novice users when interacting with the HRP-2 robot via speech commands. As the societal impact of robotic agents in future societies can only be roughly predicted, it is of clear interest how novice users themselves imagine their usage of robotic systems. In social sciences the validation of a methodological mix is typically conducted by performing a pre-study with a limited number of participants. These pre-studies can help understand limitations and shortcomings of the method as well as possible influencing factors of an experimental setting.

The pre-study is presented as follows: first related work is discussed, second we present our set-up and the results from a first exploratory study, third we discuss lessons learned for the proposed evaluation set-up.

2 EVALUATING USER EXPERIENCE AND SOCIETAL IMPACT

Several evaluation studies on perception of robots in collaborative settings have been conducted. Hinds [26] e.g. investigated people's perception of robots as working partners. She investigated differences in reliability and feeling of responsibility on the part of the human and the robot. Results showed that participants relied more on humans as working partners, but interestingly the results were different for the concept of responsibility. There was slightly no difference between participants' feelings of responsibility for a human and a robotic working colleague. Also the comparison between human-like and machine-like robots in terms of the willingness to rely on was not significant.

Another interesting aspect of the integration of robots into working environments was revealed by [3]. Based on a long-term ethnographic study, where a delivery robot was introduced into the everyday working environment of a hospital they could show that perception of the robot depends on the user group. Their results showed significant differences on how people responded to the robots in two different organization units in the hospital. The delivery robot was differently integrated into the work-flow of the two units and also the way of interaction differed.

Some aspects of UX in HRI are extracted from avatar research since many research objectives from UX in HRI are also pursued by scientists of ECAR (Embodied Conversational Agent Research) (e.g. the robot serves as a representation of a human, or robots are used to investigate embodied models of social behaviour [28]). The embodiment (or morphology) [28] has an impact on how a robot is perceived. It also implies capabilities of a system, which in some cases cannot be fulfilled. If a robot is designed biologically inspired with anthropomorphic features, then users will expect human-like capabilities. In some cases users may experience fear as robots are designed too realistic. It is important for people that there is a clear difference between humans and robots [27]. In 1970 Masahiro Mori, proposed the well known hypothetical graph of the uncanny valley, which predicted that the more human a robot looks, the more familiar it is, until a point is reached at which subtle imperfections make the robot seem eerie [15].

By incorporating existent navigation technique metaphors in HRI, a more positive UX is obtained. This can be achieved by transferring navigation strategies and interface designs of computer games (first person shooters) to robot navigation [25].

However, the evaluation of user experiences while interacting with a robot is only sparsely reflected in the current literature. Syrdal et al. investigated how context factors in HRI video prototyping influence the user's experience and attitude, by means of qualitative interviews [7]. However, to our knowledge there is no work directly referring to the evaluation of UX in terms of pragmatic and hedonic quality. Other evaluations refer to factors that are related to the concept of UX, like acceptance, see e.g. [24][21][23][22]. Attitudes [18], especially negative attitudes [29]; [5], have been investigated in detail. An excellent overview on existing measures for single UX factors in HRI can be found in [4]. However, studies on "general" user experience have not been conducted so far.

Societal impact on the other hand was assessed with a variety of methods. Sakamoto and Ono [8] used a lab-based approach to investigate if in the future robots will construct or collapse human relations. They conducted an experiment based on a communication situation between two humans and one robot with three conditions: (1) robot agrees with humans, (2) robot disagrees with humans, (3) robot agrees only with one human. However, the results of this one laboratory-based experiment could not show a clear tendency if human relations will collapse or not in future societies due to the introduction of robots.

Large questionnaire based surveys have been conducted to get insights on people's expectations towards the future society and the societal impact of robots. Arras and Cerqui [16] conducted a paper and pencil based survey with 2042 participants in conjunction with the "Robotics" pavilion at the Swiss National Exhibition Expo.02. The results testified a positive attitude towards potential robotic co-workers, robots as flat-mates, and even for robotic body parts. The authors stress on the one hand that context-related surveys with a sufficient number of participants give valuable insights on societal trends, but on the other hand they admit that the conducted survey is only valid for Europe and should be conducted in the USA and Asia as well to get more generalizable results.

The relationship of usability, user experience, (social) acceptance and societal impact is currently investigated in the Robot@CWE project funded by the 6th EU IST framework. Weiss et al. [2] showed how breaching experiments can help to understand social acceptance and societal impact factors. The following case study lays the basis for the evaluation of user experience and societal impact.

3. THE CASE STUDY

3.1 Methodological Considerations

The case study was based on two tasks, where participants had to control the HRP-2 robot via speech commands to pick up (task 1) and put down an object on another place (task 2). Four participants took part in this preliminary case study (1 female, 3 males). The youngest participant was aged 23, the oldest one 39, with an average age of 27.5 years.

The content-related research questions of this user study were:

- How do novice users experience the interaction with the humanoid robot HRP-2 when interacting via speech commands?
- How do users perceive the system in terms of usability?
- Does the general attitude towards robotics change because of the interaction with the robot?
- How do people imagine the future society after interacting with the robot?

From a methodological view point the focus of interest was to verify if the standardized questionnaires from other disciplines, as well as the first self-developed questionnaire on UX indicators are reasonable to evaluate the proposed factors. Clearly a statistical validation of the questionnaire data was not possible with four participants, but the combination of different questionnaires and the qualitative data showed tendencies which study settings and instruments can be reasonably used. Furthermore, it was necessary to investigate the workload of such a study setting for the participants and which methodological considerations have to be taken into account for future experiments.

3.2 Study Setting

The user study with HRP-2 took place at the LAAS - CNRS institute in Toulouse from 8th to 11th of July 2008. It was based on two tasks which participants had to conduct with the humanoid HRP-2 robot. 15 clones of the HRP-2 robot exist. University of Tokyo and AIST conduct research with these robotic platforms on robot autonomy. Currently the Joint French-

Japanese Robotics Laboratory (JRL) in collaboration with LAAS-CNRS and AIST are working on enabling HRP-2 understanding and executing commands like "give me the orange ball", as autonomously as possible (details on technical constraints of this work can found in [13]). This built the basis for the presented case study, which was conducted with HRP-2 number 14.

The study design was based on two tasks (see table 1) participants had to complete by giving HRP-2 speech commands More details about the implementation of the two tasks can be found in [9]

Table 1: The two Tasks

Task 1	Task 2
Instruction: "Tell the robot to pick up the orange ball. Therefore, you have to use the following commands (after each command, wait for the reaction/answer of the robot, before you start with the next one)"	Instruction: "Your task is now to tell the robot to put the orange ball on the table. Therefore, you have the following commands"
Commands needed: "14" (is the general command which activates the robot, and its name) "Go to the green box" "Look down" "Take the orange ball"	Commands needed: "14" (is the general command which activates the robot, and its name) "Turn to the left" "Go to the yellow table" "Look down" "Put the orange ball on the yellow table"
Additional commands: Turn to the left Turn to the right	Additional commands: Turn to the left Turn to the right

The tasks were introduced to the participant by the experimenter reading aloud the following scenario:

``Imagine you are working at a construction site and need a special tool. A humanoid robot, its name is fourteen, is supporting you by providing you with your needed tools. You can control the robot with predefined commands, therefore it picks-up and transports the tool''

Afterwards the participant received a microphone to give the robot the speech commands. HRP-2 executed the commands autonomously, therefore the additional commands were necessary, as it could happen that due to wrong speech recognition the robot executed motions that were not expected by the participant (e.g., the robot picks up the purple ball instead of the orange one).

Figure 1 shows the setting, with the HRP-2 robot (the execution of the motion was observed by a researcher in the background due to safety reasons), the participant with the microphone, the experimenter taking observational notes, and the green box, with a purple and an orange ball.

Figure 1: Study Setting

Four participants took part in the case study, one was female three were male and they were between 20 and 40 years old. The requirements to select the participants were: (1) know English well (the study was conducted in English); (2) not have any pre-experience with robots (as insights should be gathered how novice users experience the interaction)

3.3 Instruments and Measurements

3.3.1 Usability Measurements

The whole user study was video-taped. Task completion and task duration were recorded during the user study to address the usability indicators effectiveness and efficiency. Furthermore, a "retrospective think aloud" was conducted. After each task, participants were asked about improvements and changes they would suggest to make the interaction with HRP-2 easier in future.

Additionally the system usability scale questionnaire (SUS) was used to address the perceived usability of the interaction. This questionnaire was developed at Digital Equipment Corp. as a tool which allows a simple and quick standardized evaluation of system usability. The scale consists of ten statement-based items, giving a global view of subjective assessments of usability. The statements are rated by the participants on a five-point Likert scale from "strongly disagree" to "strongly agree". The overall scale takes into account the effectiveness, efficiency, and satisfaction of participants with a system. Thus the result of the questionnaire is a single number representing a composite measure of the overall usability of the system being studied, meaning that scores for individual items are not meaningful on their own. The range for this cumulative score is in between 0-100: "80-100: users like the system"; "60-79: users accept the system"; "0-59: users do not like the system".

3.3.2 User Experience Measurements

To get general insight on UX when interacting with the HRP-2 robot, the validated standardized AttrakDiff questionnaire was used. This questionnaire was designed to measure beyond the usability (pragmatic quality) also the user experience (hedonic quality) of a system [20]. The questionnaire is a semantic differential consisting of numerous antithetic word-pairs, e.g. "disagreeable – likable". All items have to be graduated by the participants on a scale ranging from the negative word pole to the positive word pole; the negative pole (in sense of the scale) is coded with -3 and the positive with 3.

The results are presented in four different sub scales: "PQ = pragmatic quality of the system", "HQ-I = Hedonic Quality Identification", "HQ-S = Hedonic Quality Stimulation", and "ATT = Attractiveness". The pragmatic quality describes the usability of the system and indicates how successful users are in achieving their goals using the product. The hedonic quality - identity scale describes to what extent the system allows the user to identify with it, whereas the hedonic quality - stimulation is an indicator to what extent the system can support stimulating functions, contents, and interaction- and presentation-styles. Finally attractiveness describes a global value of how the quality of the system is perceived by participants. The hedonic and pragmatic qualities are independent of one another, and contribute equally to the rating of attractiveness.

Furthermore, a questionnaire was developed (subsequently called UX-questionnaire) which tried to address five predefined UX factors with five statements each and had to be rated on a 7-point Likert scale (where 7 was the best rating in sense of the scale) by the participants.

The five factors were chosen based on an extensive literature review: Emotion [28], Embodiment [28], Feeling of Security [12], Human-oriented Perception [28], Co-Experience [14] and are part of the USUS evaluation framework [1]. Table 2 presents all items of the questionnaire with the corresponding factors.

Table 2

Statements addressing User Experience Factors

Statement	Factor*
I liked the size of the robot.	Emb
Interacting with the robot is fun.	E
When talking to the robot, I feel like talking to a human.	Co
I am happy when the robot understands my commands.	E
I think that the robot is vulnerable to hackers.	FoS
I can interact with the robot like I interact with other humans.	Co
I am disappointed if the robot does not understand my commands.	E
I liked that the robot looked similar to a human.	Emb
I hesitate to use the robot for fear of making errors that will harm me.	FoS
When working with the robot I perceive it as working in a team.	Co
I perceive the robot as a social actor	HoP
I liked that the robot has human like features: face, ears, eyes, etc.	Emb
I feat to use the robot, as an error might harm the robot.	FoS
I liked that the robot detected my face.	HoP
I feel good when interacting with the robot.	Co
I liked the physical co-location of the robot.	Emb
I perceive that the robot is intelligent	HoP
I am angry if the robot does not understand my commands	E
I liked the design of the robot.	Emb
I enjoyed talking with the robot	HoP
The robot could become a companion for me.	Co
I feel secure when working with the robot.	FoS
I felt afraid of the robot.	E
I liked that the robot understands my voice commands	HoP
I perceive the robot as safe.	FoS

*Emb: Embodiment; E: Emotion; Co: Co-Experience; FoS; Feeling of Security; HoP: Human-oriented Perception;

To measure if the participant's attitude towards humanoid robots as working colleagues changes during the user study the NARS questionnaire was used.

This questionnaire is based on a psychological scale to measure the negative attitudes of humans against robots and is originally developed by [29]. In the described user study the questionnaire was used in an English version [5]. This questionnaire tries to visualize which factors prevent individuals from interacting with robots. The questionnaire consists of 14 questions, which have to be rated on a 5 point Likert scale ranging from "Strongly Disagree" to "Strongly Agree". The 14 questions build three sub scales: S1 = Negative attitude toward situations of interaction with robots; S2 = negative attitude toward social influence of robots; S3 = negative attitude toward emotions in interaction with robots.

3.3.3 Societal Impact Measurements

To address participants' assessment of the future society regarding humanoid robotic working colleagues, an in-depth interview was conducted. The interview was based on four questions which were discussed with the participants.

- In which way could robots be integrated into working life in future?
- How could life change if robots are integrated into a construction site?
- How will the usage of robots in the working context influence the future education system?
- How could you imagine that society will support the use of robots in the future?

3.4 Procedure

The study started in a room separated from the test-setting, where participants had to fill in the NARS at first. Then the two tasks were conducted. After each task participants were asked what they thought and felt during the interaction with the robot, if they had any problems during the interaction, and if they would like to change or improve something to make the task fulfillment easier. Then participants rated the difficulty of the task. After the two tasks were carried out, participants filled in the questionnaires in the following order:

SUS

AttrakDiff

UX-Questionnaire

NARS (second round)

After the questionnaires were filled in the user study was concluded with the in-depth interview on participants' imagination about how future society could look like regarding the integration of robotic technology.

5 RESULTS

5.1Insights on Perceived Usability

Table 3 gives an overview on the task completion of all participants for all tasks. The results show that participants were very successful in solving the tasks, and thus did not experience the interaction with HRP-2 as very difficult.

Table 3: Task Completion

	Task 1	Task 2
Solved successfully	2	3
Solved with help	2	1
Not solved	0	0
Average rating of difficulty[4]	1.5	1.33

The tables 4 and 5 give an overview on the clustered comments participants stated in the retrospective think aloud. The number in brackets is the number of the participant (not a count!).

Table 4: Retrospective think aloud – Task 1

Participants liked	Participants disliked
Robot answered in an nice way (3)	Speed of the robots reaction (1, 2, 3, 4) (too slow)
No need of additional and repeated commands (3)	TP felt insecure if the robot is close enough to grasp the ball and therefore already had to look on how to proceed with the next command. (3)
Robot describes all moves very good and helpful (3)	There was orange and purple misunderstanding (1)
Impressed, really impressed, if it was doing something wrong it was my fault (2)	I did not understand what it said, I had just to give commands (1)
Amazing, like in a movie (3)	Not every move needs a command (4)
The fears the robot would not understand me, were not realized.	When it still needs human it is not useful (4)
Feeling very close to the process (1)	

Table 5: Retrospective think aloud – Task 2

Participants liked	Participants disliked
Easy to use (2, 3)	Difficult: I had to figure out and understand how it works in term machine (1)
I thought I need to talk slow, I would not have thought it is so precise (2)	I saw the robot as a machine, not an human (1)
Cool and easy to do (3)	The robot is not human like, I do not remember that the robot detected my face (4)
I was comfortable when the robot explained why he did not do the command (3)	Comprehension should be better (orange-purple) (2)
	I thought he will remind the yellow table that would be more logical (3)
	It should remind things then it gets easier (3)
	I was a little bit stressed as it did not understand the first command.
	Its movement is slow (4)
	The time between answer and movement is too long (4)
	Improve: having different programs, one for talking, one for movement (4)
	Lack of pronunciation (p of put, and p of purple)

[4] from 1 "very easy" to 5 "very difficult"

The SUS questionnaire equaled a surprisingly high result of 78.8, meaning that the participants accepted the HRP-2 robot in terms of its usability. However the SUS is no absolute ratio, it depends on the experimental context; in this case study participants evaluated the speech interaction with the robot.

5.2 Insights on User Experience

The results for the AttrakDiff questionnaire (see table 6) presented a neutral image of the general UX when interacting with the robot. The comparison of the mean values for each scale shows that people experienced the robot rather neutral (all mean values are around 0). This means that there is still room for improvement of the user experience of the robot in terms of "Pragmatic Quality", "Hedonic Quality Identification", "Hedonic Quality Stimulation", and "Attractiveness".

Table 6: Results AttrakDiff

	PQ	HQI	HQS	ATT
Mean	0.0000	-.5357	.0357	.0714
S.D.	.37796	.44224	.62133	.24744

For the UX-Questionnaire the UX factors were calculated by summing up the ratings of the participants (see table 7). They can get values from 1 to 7. The higher the value is, the more positively the participants connected the form of experience with the tested robot.

Table 7: Results UX Questionnaire

UX-Factors	min	max	mean	SD
Co-experience	3.20	4.40	3.7000	.25166
Embodiment	4.60	6.20	5.3333	.46667
Emotion	5.20	6.60	5.8500	.29861
Human-Oriented Perception	4.20	5.80	4.9333	.46667
Feeling of Security	4.20	5.80	5.2000	.38297

The factor "emotion" comes off very well with an average value of 5.85. Also "embodiment", "feeling of security" and "human-oriented perception" have quite good values with 5.33, 4.93 and 5.2. Only located in the middle of the field is the factor "co-experience" with an average value of 3.7.

The NARS questionnaire revealed to most impressive result (see table8). There is a significant difference between the rating of "negative attitude toward situations of the interactions with robots" (S 1) before and after the test. After the experiment/test the rating of the "negative attitude toward situations of the interactions with robots" is significantly lower than before the experiment. Also the rating of "Negative Attitude toward Social Influence of Robots" (S 2) has significantly changed. Before the test the rating of "Negative Attitude toward Social Influence of Robots" has been significantly higher than after the test. The rating of "Negative Attitude toward Emotions in Interaction with robots" has also decreased, but not significantly.

Table 8: Results NARS Questionnaire

		Mean	SD	t	sig.
Nars S1	t1	11.75	4.11	5.196	.014
	t2	10.25	3.95		
Nars S2	t1	14.00	2.16	3.434	.041
	t2	10.75	2.36		
Nars S3	t1	9.75	2.06	-1.507	.229
	t2	7.50	2.38		

5.3 Insights on Perceived Societal Impact

When discussing about the question in which way robots could be integrated into future working life, 3 out of 4 participants use the term "replacement", which implicates that if robots are integrated into working life, automatically something or someone other is excluded from work (e.g. humans, machines). Therefore "replacement" is a negative term.

Basically the participants are of the opinion that robots can replace on the one hand "slavery work" and on the other hand dangerous work. The term "slavery work" is defined as both repetitive tasks in construction factories and domestic work like cleaning, cooking, washing the dishes, cutting the grass or even driving a car. As examples for dangerous work the participants name "finding lost people in the mountains" or "lifting heavy things". In the participants` opinion the advantage of using robots is that they can be replaced when they are broken, the disadvantage is that human life is nowadays cheaper than robots (e.g. in China).

The following citations show that the participants are thinking differently about the integration of robots into the working life.

- "It is impressive, but dangerous, we are a human society not a robot society, we have to keep humanity" (participant 2)
- "Environments working harmful for humans" (participant 1)
- "Robots are useful as assistance" (participant 3)
- "You can replace a robot if it is broken" (participant 4)

Two participants foreground the dangers aroused through the integration of robots into working life (e.g. unemployment), the other two participants do not mention dangers but give priorities to the advantages that could accrue by the usage of robots in the working life. So half of the participants vision the integration of robots into the working life positively, half of the participants negatively.

Concerning the question how life could change if robots are integrated into a construction area, the participants forecast both positive and negative effects. One negative effect is the unemployment as a result of the replacement of humans by robots. Also negatively seen is the competition between robots and humans. The participants are of the opinion that robots will compete because they do not need to eat or drink and therefore are more reliable, meaning that the participants think that the costs of the workers are the most crucial employment-factor. It is seen positively that there will be less accidents in future working life with robots, so work will be safer, there will be less health problems and people will have more spare time where they have the possibility to do funny things while robots are doing their work (e.g. their household). Also positively noted is that work that only can be done by humans will be more valued in a future working life with robots.

Subsequently it can be said that participants think there will be more unemployed people, because robots are doing their work. At the same time people will have more time for doing funny things. What the participants do not mention is the inconsistence between increasing unemployment and increasing fun in the spare time. There will be probably increasing fun for one group of people and less fun for the other unemployed group that cannot afford robots for domestic work.

Concerning the question how usage of robots in the working context will influence the future education system, participants have two different concepts. One part of participants is of the opinion that people have to learn the usage of robots in school (e.g. school subject "technology" beginning at the age of about 15). The other part of participants thinks that children will be able to learn outside school how to use robots (e.g. "parents will teach the children the dangers of robots, it´s like with cars").

Therefore, we can distinguish between two robot-education-types:

- Type 1: A robot is only one more technology, and there is no need for special education. Computers in education are enough.
- Type 2: There should be an introduction of robots in school, so people get to know how to interact with a robot.

The last question was about how the society would support the use of robots in the future. For participants the usage of robots in future is mostly a benefit-cost-question. They think it depends on the price, the tasks the robot is able to do, the speed it is fulfilling the tasks with and so on. But also the ease of use and the main aim will be important factors. (If robots are only dedicated to economic growth they will not be as accepted as if robots that are dedicated to our life). Participants are of the opinion that if people get used to robots (if they are comfortable), they will be just another technology, like cars. For the young generation robots could be domestics, like computers, especially if they get cheaper and quicker.

6 LESSONS LEARNED FOR THE PROPOSED EVALUATION SET-UP

Considering our experiences we recognized the following issues as being crucial to successfully address UX and societal impact in a user study with a humanoid robot:

1. Participant sampling and requirements: When conducting a user study in HRI with novice users it can be valuable to consider the educational background. In the retrospective think aloud and in the societal impact interview participants often underlined their assumptions based on their education and work background: “As I am working in education; I can tell you that robots should not teach children”; “As I studied computer science, I have a mental model how the robot vision works”.
2. Method mix: It is recommended to combine qualitative and quantitative methods to gain a more holistic picture of UX. Linking the data from the pre-structured questionnaires with the retrospective think aloud and the societal impact interviews facilitated the clustering and interpretation of the gathered data. Furthermore we could show that the proposed HCI questionnaires can be fruitfully used to understand user’s experiences when interacting with a humanoid robot.
3. Study setting: The UX factor feeling of security could not be reasonably addressed in the pre-study setting where a researcher had to observe the motions of the robot. Three participants stated difficulties in answering these questions of the UX-questionnaire: “I think the robot cannot be safe, as you have to watch it all the time”. Thus, it is recommended to assess the factor feeling of security in controlled settings that do not bias the answer behavior of the participants.
4. Procedure: All participants mentioned positively that they had the possibility to fill in the NARS questionnaire a second time: “That’s good, because my attitude actually changed”.

7 CONCLUSIONS AND FUTURE WORK

The main goal of this case study was to explore how novice users experience the collaboration with the humanoid robot HRP-2 when interacting via speech commands and if the general attitude towards robotics changes because of the interaction with the robot.

To investigate the feasibility of the study set-up and the methodological approach, we conducted a first case study with four participants. Our intention was to get a better understanding of first time user reactions on humanoid robots and to find out if these reactions influence participants’ attitude towards robots. The context of a user study where participants can directly control the robot via speech commands in combination with several questionnaires and an in-depth interview have proven its value, as different UX factors of the participants could be addressed and also discussed on the reflective level with the participants (retrospective think aloud and final interview). Furthermore, the workload seemed reasonable for the participants and built the basis for three similar future experiments.

To validate the questionnaire which was based on five UX factors in term of the scales, a summative analysis of the data gathered in this case study combined with the three future user studies and a broad online survey (we are expecting a total of 500 participants) is planned.

ACKNOWLEDGEMENTS

This case study was conducted in the framework of the ROBOT@CWE project (funded by FP6-2005- IST-5). The authors like to thank all partners from the project and gratefully acknowledge the collaboration with the researchers from LAAS-CNRS, with special thanks to Anthony Mallet.

REFERENCES

[1] A. Weiss, R. Bernhaupt, M. Lankes, M. Tscheligi, ‘The usus evaluation framework for human-robot interaction’, *in AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction* (2009).

[2] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kuehnlenz, and M. A Buss, ‘Methodological variation for acceptance evaluation of human-robot interaction in public places’, in *IEEE RO-MAN 2008: International Symposium on Robot and Human Interactive Communication, Munich, Germany*, (2008).

[3] B. Mutlu and J. Forlizzi, ‘Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction’, in *HRI ’08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 287–294, New York, NY, USA, (2008). ACM

[4] C. Bartneck, E. Croft, and D. Kulic, ‘Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots’, International Journal of Social Robotics, (2009).

[5] C. Bartneck, T.Nomura, T. Kanda, T. Suzuki, and K. Kato, ‘A cross-cultural study on attitudes towards robots’, in *HCI International*, Las Vegas, USA, (2005).

[6] D. Knemeyer and E. Svoboda, ‘User experience - ux’, *interactiondesign. org*, (2005).

[7] D. S Syrdal, N. Otero, and K. Dautenhahn, ‘Video prototyping in human-robot interaction: results from a qualitative study’ *15th European Conference on Cognitive Ergonomics: the Ergonomics of Cool interaction*, ACM, New York, NY, 1-8 (2008).

[8] D. Sakamoto, and T. Ono, 2006. Sociality of robots: do robots construct or collapse human relations?. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot interaction*

(Salt Lake City, Utah, USA, March 02 - 03, 2006). HRI '06. ACM, New York, NY, 355-356.

[9] E. Yoshida, A. Mallet, F. Lamiraux, O. Kanoun, O. Stasse, M. Poirier, PF. Dominey, J-P. Laumond, and K. Yokoi, "'give me the purple ball" – he said to HRP-2 no.14', in Proc. 2006 IEEE-RAS Int. Conf. on Humanoid Robots, (2007).

[10] J. McCarthy and P. Wright 'Technology as Experience. London : MIT Press (2004).

[11] K. Battarbee, 'Defining co-experience', in DPPI '03: Proceedings of the 2003 international conference on Designing pleasurable products and interfaces, pp. 109–113, New York, NY, USA, (2003). ACM.

[12] K. Dautenhahn, M. L. Walters, S. Woods, K. L. Koay, E.A. Nehaniv, C. L. Sisbot, R. Alami, and T. Simeon, 'How may I serve you?: a robot companion approaching a seated person in a helping context', in *HRI '06: Conference on Human-Robot Interaction*, pp. 172–179, Salt Lake City, Utah, USA, (3 2006). ACM.

[13] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, K. Akachi M. Hirata, and T. Isozumi, 'The humanoid robot HRP-2', in Proc. 2004 IEEE Int. Conf. on Robotics and Automation, pp. 1083–1090, (2004).

[14] K. Severinson-Eklundh, Anders Green, and Helge Hüttenrauch, 'Social and collaborative aspects of interaction with a service robot', *Robotics and Autonomous Systems*, **42**(3-4), 223–234, (2003).

[15] K.F. MacDorman, and H. Ishiguro, 'The uncanny advantage of using androids in social and cognitive science research'. Interaction Studies, 7, 3 (2006), 297-337.

[16] K.O. Arras and D. Cerqui, 'Do we want to share our lives and bodies withrobots? a 2000 people survey', Technical report, (2005 2005). Technical report Nr. 0605-001.

[17] L. Alben, 'Quality of experience: defining the criteria for effective interaction design', *interactions*, **3**(3), 11–15, (1996).

[18] L. Takayama, W. Ju, and C. Nass, 'Beyond dirty, dangerous and dull: what everyday people think robots should do', in *HRI '08: 3rd ACM/IEEE international conference on Human robot interaction*, pp. 25–32, New York, NY, USA, (2008). ACM.

[19] M. Hassenzahl and N, Tractinsky, 'User Experience – a research agenda', Behavior and Information Technology, **25**(2), 91-97 (2006).

[20] M. Hassenzahl, 'The thing and I: understanding the relationship between user and product', in *Funology. From Usability to Enjoyment*, eds., M. Blythe, C. Overbeeke, A. F. Monk, and P. C. Wright, 31–42, Kluwer, Dordrecht, (2003).

[21] M. Heerink, and B. Kröse, B. Wielinga, and V. Evers, 'Studying the acceptance of a robotic agent by elderly users', *International Journal of Assistive Robotics and Mechatronics*, **7**(3), 25–35, (9 2006).

[22] M. Heerink, B. Kröse, B. Wielinga, and V. Evers, 'Enjoyment intention to use and actual use of a conversational robot by elderly people', in *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 113–120, New York, NY, USA, (2008). ACM.

[23] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of a robot's social abilities on acceptance by elderly users', in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pp. 521–526, (2006).

[24] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, 'The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users', in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pp. 695–700, (2008).

[25] M. W., Kadous, Sheh, R. K., and C. Sammut, 'Effective user interface design for rescue robotics', HRI '06. 1st ACM SIGCHI/SIGART Conference on Human-Robot interaction (Salt Lake City, Utah, USA, March 02 - 03, 2006), pp. 250-257, New York, NY, (2006). ACM.

[26] P. J. Hinds, T. L. Roberts, and H. Jones, 'Whose job is it anyway? a study of human-robot interaction in a collaborative task', *Human-Computer Interaction*, **19**(1-2), 151–181, (2004).

[27] S. Turkle, 'The second self: Computers and the human spirit'. Simon & Schuster, New York, (1984).

[28] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**, 143–166, (3 2003).

[29] T. Nomura, T. Kanda, and T. Suzuki, 'Experimental investigation into influence of negative attitudes toward robots on human-robot interaction', *AI Soc.*, **20**(2), 138–150, (2006).