EVALUATING THE BEHAVIOUR OF DOMESTIC ROBOTS USING VIDEO-BASED STUDIES

MICHAEL L. WALTERS, MANJA LOHSE, MARC HANHEIDE, BRITTA WREDE, KHENG LEE KOAY, DAG SVERRE SYRDAL, ANDERS GREEN, HELGE HÜTTENRAUCH, KERSTIN DAUTENHAHN, GERHARD SAGERER, and KERSTIN SEVERINSON-EKLUND.

M. *L. Walters, K. L. Koay, D. S. Syrdal, and K. Dautenhahn are with University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, UK . E-mail: {m.l.walters; k.l.koay; d.s.syrdal; k.dautenhehn} @herts.ac.uk. M. Lohse, M. Hanheide, B. Wrede, and G. Sagerer are with the Applied Informatics, Bielefeld University, 33619 Bielefeld, Germany. E-mail: {mlohse; mhanheid; bwrede; sagerer} @techfak.uni-bielefeld.de). A. Green, H. Hüttenrauch, and K. Severinson-Eklundh are with KTH Royal Institute of Technology School for Computer Science and Communication (CSC), SE-100 44, Stockholm, Sweden. E-mail: {green; hehu} @csc.kth.se and kse@kth.se)*

Abstract

As robots are used for more complex applications in human oriented environments, they will have to be able to interact with inexperienced users. Findings from human-human interaction and human-computer interaction research are relevant, but often limited because robot are different from both humans and computers . Therefore, new methods have to be developed in Human-Robot Interaction (HRI) research to inform the design of these robots in order to build robots that are suitable for inexperienced users. A Video-based HRI (VHRI) methodology was used to carry out a multi-national HRI user study which employed our robot BIRON (BIelefeld RObot companioN) which is designed for use in domestic environments. Participants watched videos of the robot system interacting with a human actor, and rated two different robot behaviours (extrovert and introvert). Their perceptions and ratings of the robot behaviours differed with regard to verbal output and person following by the robot. Participants' ratings of the robot behaviours were evaluated and compared and suggestions made with the aim of improving the human-robot interaction.

Keywords: HRI, User-trial, Video. Methodology, Multi-national

1. INTRODUCTION

Robots are now finding uses as toys (e.g. Aibo [1], Furby [2], Lego Mindstorms [3]), for cleaning (e.g. Roomba [4]) and healthcare (e.g. Paro [5]) and inexperienced users are expected to interact and control these machines. These robots are expected to fulfil their tasks effectively and users should also like to interact with the systems and want to use them for a long time. Whenever technical devices are built for novice users, design decisions should be informed and evaluated with the help of potential users. User studies should be part of the development cycle providing useful hints for improvement both of the technical and interaction capabilities of the robot.

Previously, Human-Robot Interaction (HRI) user studies have typically employed live human-robot experiments in which humans and real robots interact in various experimentally controlled scenarios (e.g.. Green et al. [6] and Walters et al. [7]). Live HRI trials are generally complicated and expensive to run and typically test a relatively small sample of possible users. An evaluation approach originally proposed by [8] is a Video-based HRI (VHRI) methodology for user studies, which can provide a supporting method to live HRI user trials. In VHRI studies, interactive robot behaviours are recorded on videotape, which is then shown to many viewers who are then asked to rate the behaviours they watch. The method enables researchers to conduct studies with a large sample of participants in a relatively short time. We chose it as a pilot method to evaluate the domestic robot BIRON (cf. [9-11]). This paper presents the results of a study with 233 participants from Germany, Great Britain and Sweden

2. EXPERIMENTAL METHODOLOGY DEVELOPMENT

In the design of products and systems, video has long been used as valid medium for visualising, prototyping and user testing a wide range of products [12]. We have also previously developed and verified the use of VHRI as a methodology for performing HRI user trials, and the methodology chosen was adapted from that we have employed in our previous work [8, 13]. In these initial studies the VHRI method was originally developed and verified, and the results obtained from participants who viewed a video recording of another person participating in interactions with a robot were found to be comparable to those obtained from participants in the same live interactions. For more complete details on these studies see Woods et al. [14, 8], but a summary of aspects relevant to the development of the VHRI methodology is provided here.

2.1 Video and Live HRI Comparison Studies

A first set of live and video based comparative HRI trials was a pilot study that performed a limited exploratory investigation to assess the potential for the comparability of people's perceptions from live and video HRI trials. Fifteen participants took part in live HRI trials and videotaped HRI trials in which the scenario for both trials was identical, involving a robot fetching an object and carrying it to them using different approach directions. Findings from the pilot trials indicated moderate to high levels of agreement for participants' preferences and opinions for both the live and video based HRI trials.

In order to verify these pilot trial findings, and to extend the investigation, a series of video and live HRI trials were performed with a larger sample set of forty two participants, and a wider range of HRI situations involving a robot approaching human participants. In this main study, additional controlled conditions included the human participants sitting in an open space, sitting at a table, standing in an open space and standing against a wall. The subjects experienced the robot approaching from various directions for each of these contexts in HRI trials that were both live and video-based. There was a high degree of agreement between the results obtained from both the live and video based trials using the same scenarios. The main findings from both types of trial methodology were: Humans strongly did not like a direct frontal approach by a robot, especially while sitting (even at a table) or while standing with their back to a wall. An approach from the front left or front right was preferred. When standing in an open space a frontal approach was acceptable and although a rear approach was not usually the most preferred, it was generally acceptable to subjects if physically more convenient. Significant comparable results were also obtained for both sets of trials with regard to robot approach speed and distance.

Overall, the findings from these two sets of experiments supported the use of the video based HRI methodology for developing and trying out new innovative studies that are in the pilot phase of testing. Naturally, it is appreciated that there are numerous limitations of using video footage for HRI studies, and it should be appreciated that they are not a replacement for live HRI studies. It is expected that the more interaction is involved between robot and participant in a given trial, the less suitable video trials would be due to the increased importance of aspects of embodiment, dynamics and contingency of interaction. However, for the particular research questions that we considered in this current study, the contingency of robot and human movements played a less crucial role and therefore the results justify our choice of Video-based HRI trials.

3. METHOD

The practical aim of HRI as a discipline is to design robots that are as effective, efficient and usable as possible for all interaction roles the user might take. Methodologies from other fields are adapted and new ones have to be developed. Previously, mostly theories and methods from psychology and human-computer interaction (HCI) have been applied to HRI research (e.g. [15, 16]). The aim in HCI and usability research in general is to evaluate systems with potential users in realistic conditions including environments and tasks [17, 18]. Several approaches in this direction have been taken for robotics (e.g. [19, 20]), though depending on the tasks of the robot and the context it is used in, it is often difficult to conduct live user studies. The effort to run full scale HRI trials can be large and the number of participants in HRI live user studies is typically relatively small. In live HRI trials, systems might show different behaviours with different users, which may cause concerns regarding comparability of the data.

Video trials provide a complementary method to conduct studies with many participants and increased comparability. The authors in [8, 14] sum up the main advantages of video based HRI trials as follows: "1) reach larger numbers of subjects as they are quicker to administer, 2) easily incorporate subjects' ideas and views into later video trials simply by recording extra or replacement scenes into the video based scenarios, 3) carry out trials exposing groups of subjects to an HRI scenario simultaneously, 4) prototype proposed live trial scenarios to avoid wasted effort and test initial assumptions, 5) allow greater control for standardised methodologies (i.e. exactly the same robot behaviours, exact trial instructions etc.)" .

Additional compelling reasons also supported the VHRI method in the present study. Firstly, our focus was on the evaluation of the interaction and not of technical components. Video studies do not allow for a technical evaluation of the system but, nevertheless, are suitable to research user experience. Secondly, the HRI trial can be conducted at different places (in this case Bielefeld University, Germany; University of Hertfordshire, Great Britain) with many participants at a time (e.g. in a university course). The robot effectively can be brought to the subjects, something which would have been difficult to organize in live trials in different countries. In contrast, in a video study all subjects judge the same robot behaviour and the language can be dubbed. Thus, comparability between groups is very high.

4. ROBOT SYSTEM AND SCENARIO

The robot used for the trials is called BIRON (see Figure 1) and is based on a Pioneer PeopleBot platform. A Sony EVI D-31 pan-tilt colour camera is mounted on top of the robot at a height of 142 cm to acquire images of the upper body part of humans interacting with the robot and to focus referenced objects. An additional camera is used to capture hand movements in order to recognize deictic references. A pair of AKG far-field microphones is located just below the touch screen display at a height of approximately 107 cm enable BIRON to localize speakers. Finally, a SICK laser range finder mounted at the front at a height of 30 cm measures distances within a scene to detect pairs of legs and to navigate. The development of the robot BIRON was framed by a home tour scenario, which envisions household robots able to adjust to new environments like a user's home. The environments have to be explored together with the customer, who probably is a rather inexperienced user, who has to teach important objects and places to the robot.

Therefore, a home tour robot must exhibit capabilities for natural interaction including understanding of spoken utterances, co-verbal deictic reference [21], verbal output, referential feedback, and person attention and following [22]. The study presented here focused on two different robot behaviours (categorized as introvert and extrovert). Previous experiments have shown that robot personality has a major influence on HRI (e.g. [23, 24, 25, 26, 27]). Once the robot enters the home of a person its personality becomes even more important because nobody wants to live with a robot she does not like. We therefore aim at developing a range of behaviours that allow the robot system to

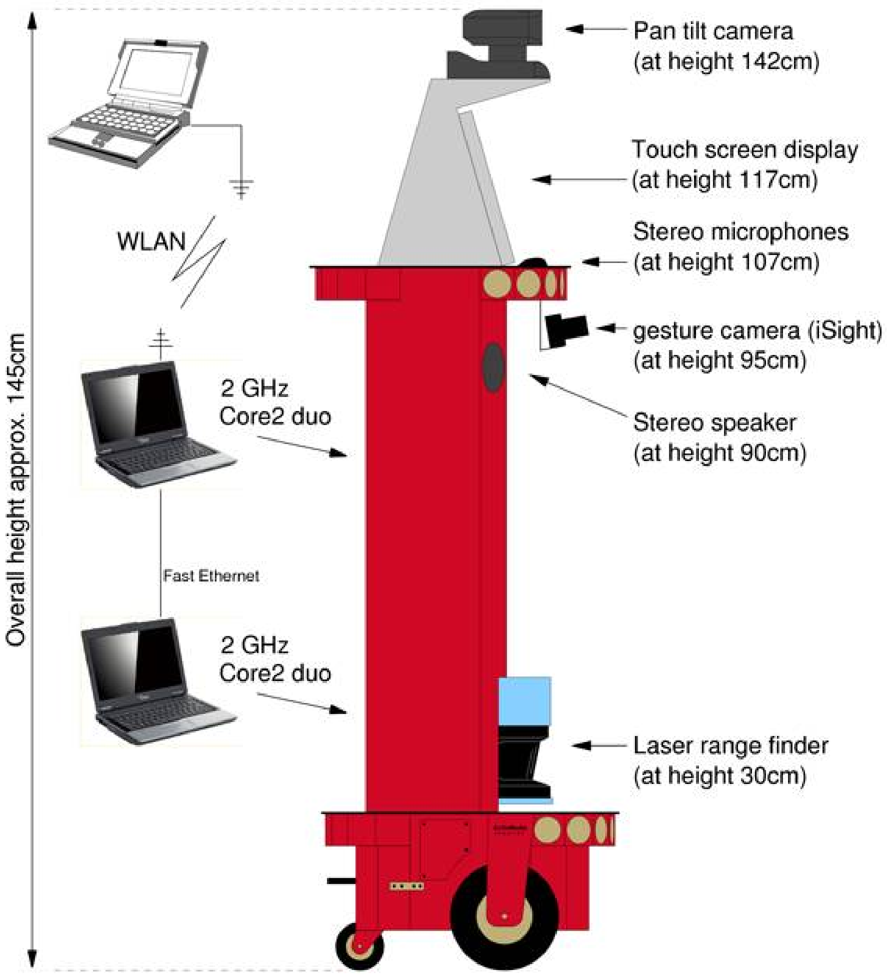

Figure 1: BIRON (BIelefeld RObot CompanioN)

adapt to the users' preferences. In the related work cited above, perception of personality was usually influenced by changing the robot appearance. As [28] found for HRI and [29] for virtual agents, speech might also influence human-machine interaction even more than appearance. We therefore developed two different interactive behaviours (labelled here, extrovert and introvert) based on an analysis of the verbal interaction. Moreover, we compared the effects of different movement patterns of the robot.

According to [30], extrovert personalities are described as sociable, friendly, talkative and outgoing. Introverts are quite introspective, and prefer to be with small groups of people. We tried to model these behaviours in the verbal behaviour of the robot and in the way it follows a person when entering a room. The main research questions addressed are:

Do subjects recognize differences between the two robot behaviours (extrovert and introvert)?

Which of the behaviours do subjects prefer?

Is the robot displaying extrovert behaviour rated as being more friendly, intelligent and/or polite than the one displaying introvert behaviour?

We do not focus on how the robot behaviours are rated by people with different personalities because we first wanted to verify whether the robot behaviours were perceived as being distinct from each other. To test this, subjects were divided into three groups, the first two watching one robot behaviour (I: introvert or E: extrovert), the last one judging both (B). Three groups were necessary to test in-group as well as inter-group differences. All participants of a group (e.g. a course) watched the videos together. Figure 2 displays the experimental procedure for each group. The whole experiment took about 25 to 30 minutes for the short conditions (I and E), and 35 to 40 minutes with the longer one (B). The development of the videos and questionnaires is described in the following section.

4.1 Videos

The main aim of the study was to test an the robot performing a task in with a human in a realistic environment. Therefore, the VHRI trial videos were shot in a real apartment (Figure 3) where the robot system was set up and functional, with an actor playing the part of the participant. The videos were made following guidelines which were based on the experience gained from running the previous VHRI trials. They specified the mix of first and third person views to be shown, the forms of editing and effects allowed, and provide a standard formula for VHRI videos to follow: An initial wide angle view of the HRI area is always shown to establish the initial spatial relationship of robot

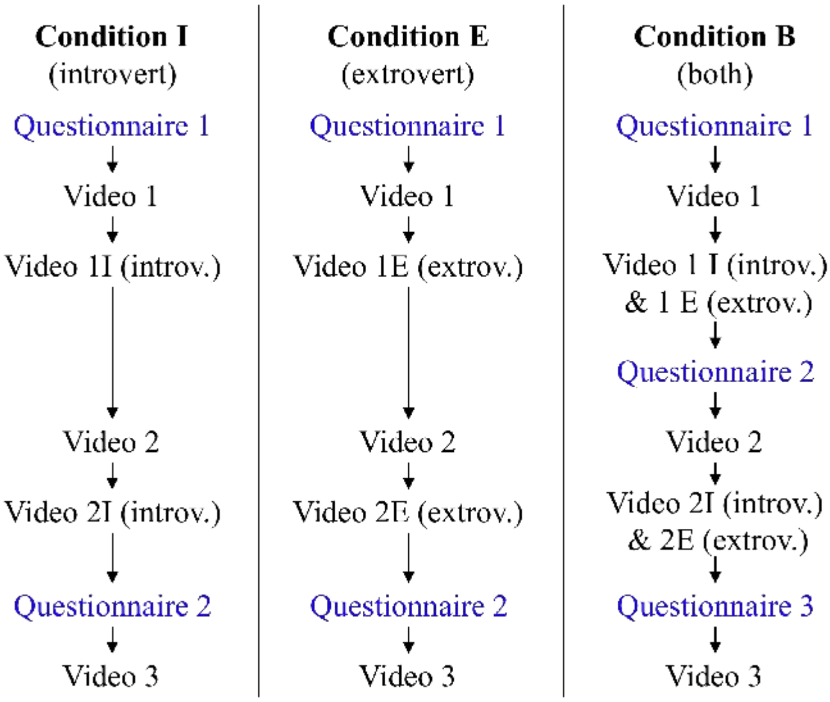

Figure 2: Experimental Procedure

and actor(s) and provide an overview of the scenario and the HRI. Then a series of first and third person views which should show the action primarily from the users point of view in order to enhance the viewers perception that they are in the middle of the action. There should be no first person views from the robot's point of view in order to reinforce the viewers empathy with the human, not the robot. All action should preferably be shown happening in "real time" with no quick cuts made from edited sequences to artificially enhance the interest of the video. Where a cut away is made to signify a passing of time, a fade out – fade in transition should employed. A subtitle to explain what was happening during the period cut by the fade transition is also acceptable.

The video was composed of three parts. Video 1 showed the introduction to the scenario with the robot being delivered and assembled by a mechanic, and incidently provides additional information about the scenario of a domestic robot which can easily be purchased, set up and employed by inexperienced users. While this video was the same for all conditions, two different robot behaviours were recorded for most of the home tour (Videos 1I & 1E and 2I & 2E).

Videos 1I and 1E presented a user (enacted by a professional actress) greeting the robot and showing objects in the living room, with the robot displaying different verbal behaviours. Video 2 was identical for all conditions and presented the robot on its way from the living room to the dining room. Videos 2I and 2E showed the user guiding the robot into the kitchen, again displaying two different behaviours. The final video (Video 3) was identical for all groups, and showed BIRON driving back to the living room autonomously. Different perspectives (first person view, third person view – see Fig. 4) were included in all videos, as recommended by VII to facilitate the viewer's comprehension.

The two robot behaviours consisted of different verbal and movement interaction patterns. Robot behaviour I was intended to be introverted and was designed to be less proactive. The robot in this condition waited until it was addressed by the user before talking. Apart from that, the robot talked little and used brief sentences which shortened the interaction significantly (Video 1I). When the user guided the system through a door into the kitchen it needed to be steered directly by commands (Video 2I).

Robot behaviour E was rather extroverted. When the actress entered the living room, the robot addressed her instead of waiting for her to start the conversation. Moreover, the extrovert BIRON was more talkative. The robot uttered longer sentences which were also more elaborate (Video 1E). In this condition, the robot entered the kitchen autonomously. It simply followed the user instead of waiting for instructions (Video 2E). The following example illustrates the difference between the extrovert and introvert verbal behaviour:

Introvert (I): User: Hello.

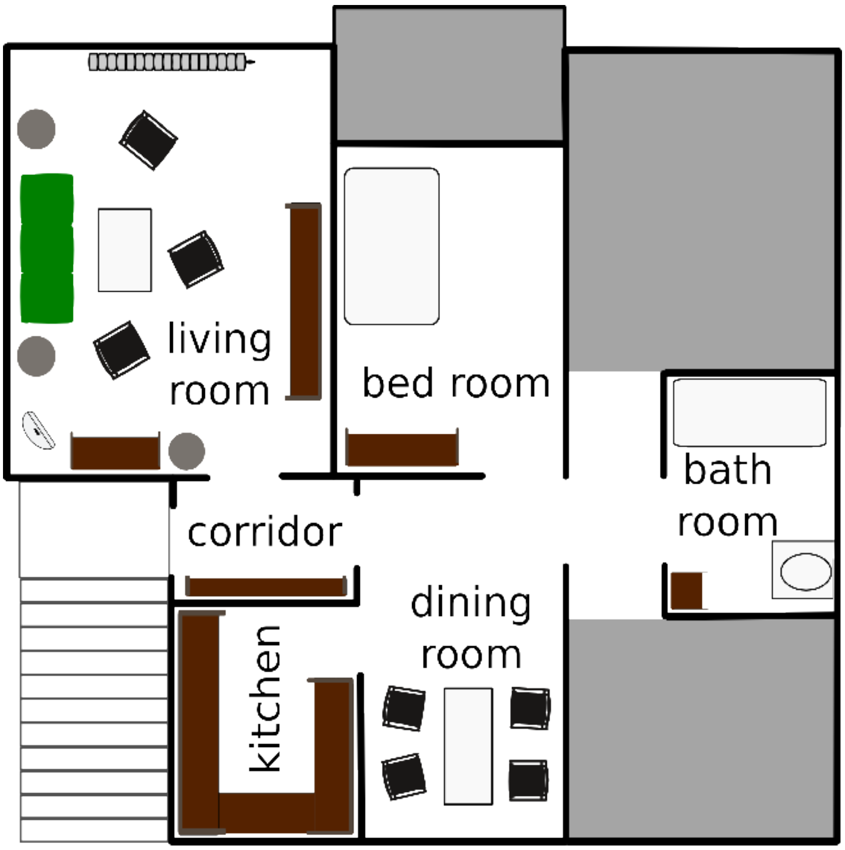

Figure 3: Robot Apartment

Figure 4: First person and third person views of the scene

Robot: Hello.

Extrovert (E): Robot: Hello. My name is BIRON. What's your name?

User: I'm Tina.

Robot: Nice to meet you Tina.

4.2 Questionnaires

Participants filled in questionnaires to rate the interactions. Before they watched the videos, all groups completed a first questionnaire, which gained basic demographic data, including age, course of study, and gender, and they also rated their experience with computers and robots. They indicated which robots they knew out of a list of 10. The participants then watched the videos.

People rating only one robot behaviour (condition I and E) watched all the videos of their condition at once, apart from the autonomous return of the robot to the living room (Video 3). Before watching this final video, they answered the second questionnaire. Subjects rating both behaviours (condition B) watched the interaction in the living room, answered the second questionnaire, watched the guiding to the kitchen and the interaction in the kitchen, answered the third questionnaire, and then finally watched the robot return to the living room. The questionnaires for this group contained the same items as the ones for the other two conditions. But participants answered a set of questions for each robot behaviour. Sequence effects in condition B (both behaviours) cannot be excluded since the videos were only shown to one group in each country. Thus, counterbalancing was not possible.

5. PRE-TEST

A pre-test was run to identify any problems in the design of the study, the questionnaires, and the videos. It was conducted in German with 54 students in three different courses. Students were divided into three groups where all three conditions were tested. The pre-test brought some insights that helped us to improve the videos and the questionnaires. An advantage identified by [14], is that single video scenes can easily be changed or replaced. After the pre-test, this was advantageous regarding shortening the overall length of the video.

Participants in the pre-test watched the robot walk back to the living room before they filled in the final questionnaire. The robot travelling back autonomously to a room previously learned is an intelligent behaviour. This turned out to overshadow the differences between the two robot behaviours. We therefore decided to have participants fill in the second questionnaire before showing the concluding part of the video. The first version of the questionnaire contained several open questions (Which robots do you know?; Name adjectives to describe the robot.). With the help of the pre-test we replaced these questions by scales to save time and to get easily comparable answers. In the pre-test subjects listed many robots they knew. From these answers the eight most frequently named robots were Aibo, Kismet, mars explorer, Asimo, soccer robot, Lego Mindstorms, Roomba, R2D2. We included these robots plus BIRON and a "service robot for the home" in the main trial questionnaire to explore whether people were familiar with the domain studied in these trials.

We also analysed the adjectives people used in the pre-test to describe the robot behaviour. Categories containing words with synonymous meaning were created and afterwards we selected a word which best described each group, and also a paired word with an opposite meaning. In the new questionnaire subjects had to rate 14 adjective pairs on a 5-point rating scale. This further increased comparability between subjects and decreased the time to answer the questionnaire. The scale consisted of adjectives which were chosen as appropriate to divide between the two behaviours tested (active, passive; interested, indifferent; talkative, quiet) and others that might result from the perception of different robot personalities (intelligent, stupid; predictable, unpredictable; consistent, inconsistent, fast, slow; polite, impolite; friendly, unfriendly; obedient, disobedient; diversified, boring; attentive, inattentive). Some other terms investigated the general usefulness of the robot (useful, useless; practical, impractical).

6. RESULTS

The results presented here include data acquired in a study with 200 participants in Germany (109) and UK (91). For logistical reasons, a study with much smaller participant numbers (33) and under a slightly different experimental procedure was run in Sweden. Therefore, the data from this limited study is considered separately to the UK and German data. All participants were assigned to one of the controlled experimental conditions (I (introvert)=62; E (extrovert)=72; B (both)=66). Their mean age was 23.95 years, 108 were male, 92 female. All German participants were students, whereas in UK 10 people belonged to the academic staff. 46.5% had a background in computer science (Germany: 30%, UK: 66%). The rest came from other disciplines (linguistics, German studies, media science, psychology, business, and health communication).

All participants had some experience working with computers (mean=3.97 on a scale of 1 (no experience at all) to 5 (a lot of experience)). However, most had little experience of interacting with robots (mean=1.65 on a scale of 1 (no experience at all) to 5 (a lot of experience)). Nevertheless, the majority indicated they knew some robots (mean=3.94 out of 10; min=0, max=10, sd=2.8), the best-known being: R2D2 (66.5%), Aibo (62.5%), mars explorer (49.5%), soccer robot (46%), and Asimo (45%). Only 14.5% knew BIRON.

Firstly, we analysed the questions "How much do you like the robot?" and "How satisfied are you with the robot's behaviour?" to find out whether subjects actually noticed a difference between the robot behaviours and if one was preferred. Table 1 presents participants' ratings. No inter-cultural differences were found in this study which supports the assumption that videos with dubbed language can be shown in various countries. It should be borne in mind that the sample was quite homogeneous as both countries were Western European. Table 1 illustrates that participants showed a significant preference for the extrovert robot behaviour (E). Both questions (How much do you like the robot? and How satisfied are you with the robot's behaviour?) were answered in favour of behaviour E. To investigate the significance of the differences, for conditions I and E (one robot behaviour) a one-way ANOVA was calculated; for condition B (both robot behaviours) a T-Test for paired samples was conducted (Likeability condition I and E: $F=21.278$; $df=1,130$; $p<.001$; condition B: $T=-8.231$; $df=64$; $p<.001$; Satisfaction with robot behaviour condition I and E: $F=5.917$; $df=1,132$; $p=.016$; condition B: $T=-8.079$; $df=64$; $p<.001$). This finding is supported by 95.2% of the subjects in condition B indicating that they noticed a difference between robot behaviours in videos 2I and 2E.

However, this does not hold true for the second rating of group B that judged both robots after the kitchen entry scene. Even though 63.5% noticed a difference between robot behaviours I and E, the ratings of the likeability and satisfaction with the robot behaviour did not differ. Reasons are given in the following analysis of the adjective ratings of the behaviours.

For the analysis of the adjectives, again a one-way ANOVA was calculated for the groups that rated one robot behaviour (Table 3), and a T-Test for the group that rated both the extrovert and introvert behaviour (Table 4). The differences between the ratings of behaviour I and E were obvious. Firstly,

Table 1: Likeability and Satisfaction with robot behaviour (Mean on a scale of 1 (very low) to 5 (very high) for conditions I, E, B (questionnaires 2 and 3))

			Both (Question 2)		Both (Question 3)	
	I N=62	E N=72	I N= 66	E	I	E
Likeability	2.46	3.27	2.20	3.18	2.33	2.29
Satisfaction	2.45	2.88	2.23	3.12	2.42	2.30

behaviour E was rated as being significantly more active, talkative, and interested. This proves that the modelling of the behaviours was successful. However, results were different for the second rating of the behaviours by group B after the kitchen entry scene. The participants could not distinguish between extrovert and introvert behaviour.

Neither door crossing was preferred. This result might partly be due to the fact that the door crossing scenes were insufficiently meaningful and the viewers of the videos did not recognize a difference. However, there might be strong preferences for one person following behaviour in live user studies where people actually might feel comfortable with the robot or not.

The analysis of the ratings of group I and E, and of group B after the interaction in the living room (questionnaire 2, Tables 2 and 3) shows that the difference between these ratings is greater when people watched both robot behaviours, and were therefore able to compare them. This might partly be due to the fact that groups I and E rated the robot only once after the kitchen entry scene. However,

Table 2: One-way Anova of ratings of Robot Behaviours in condition I and E (mean on a scale of 1 (not at all) to 5 (very much), F-value (df=1, 132), and significance)

Attribute	Mean I	Mean E	F value	Significance
Active	2.30	2.89	12.247	.001**
Talkative	2.00	2.93	26.145	<.001**
Interested	2.87	3.26	5.358	.022*
Attentive	3.54	3.59	.105	.747
Fast	1.61	2.00	6.813	.010*
Consistent	3.22	3.31	.361	.549
Predictable	3.35	3.44	.256	.614
Polite	4.05	4.31	3.149	.078
Friendly	3.56	4.03	9.218	.003**
Obedient	4.16	4.31	.894	.346
Diversified	2.12	2.69	9.218	.003**
Intelligent	2.98	3.34	4.433	.037*
Practical	2.10	2.27	1.026	.313
Useful	2.18	2.13	.087	.768

Table 3: T-Test for paired samples for Rating of Robot Behaviours in Condition B (mean on a Scale of 1 (not at all) to 5 (very much), T (df=65), and significance for ratings questionnaire 2 (questionnaire 3))

Attribute	Mean I	Mean E	T	Significance (2-tailed)
active	2.23 (2.50)	3.85 (2.48)	-12.734 (.123)	<.001** (.902)
talkative	1.97 (2.18)	4.02 (2.33)	-15.086 -(1.067)	<.001** (.290)
interested	2.41 (2.70)	3.89 (2.61)	-10.841 (.725)	<.001** (.471)
attentive	2.68 (3.09)	3.50 (2.86)	-7.083 (1.997)	<.001** (.050*)
fast	2.02 (1.89)	2.58 (1.92)	-4.511 (-.281)	<.001** (.780)
consistent	3.30 (3.20)	3.39 (2.95)	-.760 (2.898)	.450 (.005**)
predictable	3.42 (3.47)	3.06 (3.06)	2.168 (2.924)	.034* (.005**)
polite	2.98 (3.21)	4.12 (3.02)	-7.855 (1.659)	<.001** (.102)
friendly	2.88 (2.94)	4.03 (3.03)	-9.114 (-.903)	<.001** (.370)
obedient	3.55 (3.80)	3.56 (3.59)	-.136 (1.873)	.892 (.066)
diversified	1.88 (2.03)	3.02 (2.05)	-8.550 (-.155)	<.001** (.877)
intelligent	2.65 (2.58)	3.55 (2.53)	-7.421 (.382)	<.001** (.704)
practical	2.21 (2.17)	2.35 (2.06)	-1.732 (1.069)	.088 (.289)
useful	2.21 (2.12)	2.45 (2.00)	-2.248 (1.425)	.026* (.159)

the tendency of the results was the same for most items. In general, the extrovert robot behaviour was rated significantly more friendly, diversified, fast, and intelligent.

Politeness was rated very high in both conditions (I and E). Only the direct comparison of group B shows that the extrovert behaviour was judged as being significantly more polite. Altogether, the results indicate that the verbal behaviour of the robot is a powerful means to model robot personality traits. Even though a clear preference for robot behaviour E was found, the behaviour only had a small effect on the perceived usefulness of the system. In all the conditions, people did not rate the robot as being very useful or practical (see Tables 2 and 3). Reasons for this are, firstly, that BIRON did not perform any manipulative tasks in the video because the study focused on more general behaviour. Secondly, the robot used has no kind of manipulator to actually provide services in the household, such as picking up glasses for example.

The findings from the Swedish study are generally supportive of those from the UK and German studies (see Table 4). Essentially, the direction of preference is the same in the Swedish sample with some exceptions i.e. if the difference between A and B is negative in the UB/UH sample, it tends to be the same in the KTH sample). The exceptions between the two sample sets are with the attributes Attentive, Predictable, Useful,/Usable, Obedient and Consistent. However, the differences between these attributes are not significant within the UH/UB sample itself for these adjectives . However, for

Table 4: The introverted (I) and extroverted (E) mean robot attribute ratings from the limited Swedish (KTH) data compared to the combined UK (UH) and German (UB) mean attribute ratings.

Property (Swedish translation)	KTH Robot I (n=13)	KTH Robot E (n=13)	UH/UB Robot I	UH/UB Robot E	Property
Intelligent	2.6	3.0	2.98	3.34	Intelligent *
Förutsägbar	3.6	3.4	3.35	3.44	Predictable
Konsekvent	3.8	3.6	3.22	3.31	Consistent
Pratsam	3.1	4.0	2.00	2.93	Talkative**
Snabb	1.5	2.4	1.61	2.00	Fast *
Intresserad	3.4	3.9	2.87	3.26	Interested *
Aktiv	2.8	3.8	2.30	2.89	Active **
Hövlig/Artig	4.1	4.3	4.05	4.31	Polite
Vänlig	3.6	4.1	3.56	4.03	Friendly**
Lydig	4.3	4.0	4.16	4.31	Obedient
Intressant	2.6	3.0	2.12	2.69	(Diversified)**
Användbar	2.6	2.9	2.18	2.13	Usable
Uppmärksam	2.9	2.9	3.54	3.59	Attentive
Praktisk	2.5	2.9	2.10	2.27	Practical

all significant differences in the UH/UB sample, the KTH sample seems to display a similar trend.

7. CONCLUSION

The video-based HRI study methodology has been further developed and has demonstrated the main advantage of reaching many participants (200) in geographically distant places in a very short time. Across the three countries that participated in the studies, similar trends in the findings have been identified. This is probably because the three counties were all from Northern Europe and historically have had major contacts and exchanges of populations with each other, so it is likely that the people form these counties would share many attutudes.

Participants viewed videos of BIRON and rated the two different behaviours exhibited by the robot (extrovert, introvert). They noticed the differences between the behaviours and preferred the extrovert robot. Traits like intelligence, interest, friendliness, and diversity were more strongly associated with extrovert behaviour, which is also true in human-human interaction. These attributions were found mainly to be a result of the dialogue design, because they did not hold true for a door crossing (space negotiation) behaviour. However, this finding might be due to restrictions of the video based HRI method, and might change in live user studies. Nevertheless, the study gives helpful insights into users' preferences, which guide the current system design and implementation. It can be considered a powerful supplement to interactive user studies in realistic settings with working prototypes.

ACKNOWLEDGEMENT

The work described in this paper was conducted within the EU Integrated Project COGNIRON ('The Cognitive Robot Companion' - www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

REFERENCES

[1] Aibo: http://support.sony-europe.com/aibo/

[2] Furby: http://www.hasbro.com/furby

[3] Lego Mindstorms: www.lego.com/en-us/products/page2.aspx

[4] Roomba: www.roomba.ch

[5] Paro: http:/www.japan-photo.de/paro.htm

[6] A. Green, H. Huttenrauch, K. Sevderinson Eklundh, Applying the Wizard of Oz Franework to Co-Operative Service Discovery and Configuration, Proc. of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004). Okayama, Japan, pp. 575-580, (2004).

[7] M. L. Walters, S. N. Woods, K. L. Koay, K. Dautenhahn, Practical and Methodological Challenges in Designing and Conducting Human-Robot Interaction Studies, Proc. of the AISB'05 Symposium on Robot Companions, Hard Problems and Open Challenges in Human-Robot Interaction, Hatfield, UK, pp. 110-119, (2005).

[8] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, Comparing Human Robot Interaction Scenarios Using Live and Video Based Methods: Towards a Novel Methodological Approach. *Proc. of the 9th International Workshop on Advanced Motion Control, (AMC'06),* Istanbul, Turkey, pp. 750-755, (2006),

[9] J. F. Maas, T. Spexard, J. Fritsch, B. Wrede, G. Sagerer, BIRON, what's the topic? - A Multi-Modal Topic Tracker for improved Human-Robot Interaction, *Proc.* of the *IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'06),* Hatfield, UK, pp. 26-32, (2006).

[10] S. Li, B. Wrede, G. Sagerer, A dialog system for comparative user studies on robot verbal behavior, *Proc. 15th Int. Symposium on Robot and Human Interactive Communication (ROMAN'06),* Hatfield, UK, pp. 129-134, (2006).

[11] M. Lohse, K. Rohlfing, B. Wrede, and G. Sagerer, Try something else. When users change their discursive behavior in HRI. *Proc. of IEEE International Conference on Robotics and Automation (ICRA 08),* Pasadena, USA, pp. 3481-3486, (2008)

[12]S. Ylirisku, J Buur, *Designing with video*, Springer, (2007)

[13]M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. te Boekhorst and K. L. Koay, Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots* 24:2, pp. 159-178, (2008).

[14]S. N. Woods, M. L. Walters, K. L. K.oay, K. Dautenhahn, Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials, *Proc. of The 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)*. Hatfield, UK, pp. 51-58, (2006).

[15]P. H. Kahn, H. Ishiguro, B. Friedman, T. Kanda, N.G. Freier, R.L. Severson, and J. Miller, What is a human? – Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies: Social Behavior and Communication in Biological and Artificial systems*. 8(3), pp. 363-390, (2007)

[16]B. Scasselatti, Using robots to study abnormal social development. *Fifth International Workshop on Epigenetic Robotics (EpiRob).* Nara, Japan. pp. 11-14,(2005).

[17]D. J. Mayhew, *Principles and Guidelines in Software User Interface Design*. Prentice Hall, (1991).

[18]B. Shneiderman, *Designing the User Interface: Strategies for effective Human-Computer Interaction*. 3rd Edition. Addison Wesley, (2003).

[19]H.A. Yanco, J.L. Drury, and J. Scholtz, Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition. *Human-Computer Interaction* 19 (1&2), pp. 117-149, (2004).

[20]M. Lohse, M. Hanheide, A. Green, H. Hüttenrauch, B. Wrede, G. Sagerer, and K. Severinson-Eklundh, "BIRON, this is a table!" – A corpus in multimodal Human-Robot Interaction. Technical Report, University Bielefeld. (2008).

[21]A. Haasch, S. Hohenner, S. Hüwel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G. A. Fink, J. Fritsch, B. Wrede, and G. Sagerer, BIRON – The Bielefeld Robot Companion, In *Proc. Int. Workshop on Advances in Service Robotics*, eds., Prassler, E., Lawitzky, G., Fiorini, P.; and Hägele, M., Fraunhofer IRB, Stuttgart, Germany, pp. 27–32, (2004).

[22]J. Fritsch, M. Kleinehagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer, Multi-Modal Anchoring for Human-Robot-Interaction, Robotics and Autonomous Systems, *Special issue on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*. Eds. Coradeschi, S., and Saffiotti, A., Elsevier Science, vol. 43, number 2-3, pp. 133-147, (2003).

[23]R. Gockley, J. Forlizzi, and R. Simmons, Interactions with a moody robot. in HRI '06: *Proc. of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI 06)*. New York, NY,

USA, pp. 186–193, (2006).

[24]A. Tapus and M. J. Mataric´, User Personality Matching with Hands-Off Robot for Post-Stroke Rehabilitation Therapy, *Proc. of the International Symposium on Experimental Robotics (ISER-06),* Rio de Janeiro, Brazil, (2006).

[25]S. N. Woods, K. Dautenhahn, J. Schulz: Corrigendum to "Exploring the design space of robots: Children's perspectives" [Interacting with Computers 18 (2006) 1390-1418]. *Interacting with Computers* 19(5-6): p. 644, (2007).

[26]S. Woods, K. Dautenhahn, C. Kaouri, R. te Boekhorst, and K. L. Koay, "Is this robot like me? Links between human and robot personality traits," *Proc. of the IEEE-RAS International Conference on Humanoid Robots (Humanoids2005)*, Tsukuba, Japan, pp. 375- 380, (2005).

[27]B. R. Duffy. Anthropomorphism and the social robot. *Robotics and Autonomous Systems* (42), pp. 177-190, (2003).

[28]M. A. Walker, J. E. Cahn, and S. J. Whittaker, Improvising linguistic style: Social and affective bases for agent personality. In *Proc. of the First International Conference on Autonomous Agents (Agents 97), (1997).*

[29]H. J. Eysenck. *The structure of human personality*. London: Methuen, (1953).