Welcome to the Future – How Naïve Users Intuitively Address an Intelligent Robotics Apartment

Jasmin Bernotat, Birte Schiffhauer, Friederike Eyssel, Patrick Holthaus, Christian Leichsenring, Viktor Richter, Marian Pohling, Birte Carlmeyer, Norman Köster, Sebastian Meyer zu Borgsen, René Zorn, Kai Frederic Engelmann, Florian Lier, Simon Schulz, Rebecca Bröhl, Elena Seibel, Paul Hellwig, Philipp Cimiano, Franz Kummert, David Schlangen, Petra Wagner, Thomas Hermann, Sven Wachsmuth, Britta Wrede, and Sebastian Wrede

Cluster of Excellence Cognitive Interaction Technology (CITEC)
Bielefeld University, Germany
`Jasmin.bernotat@uni-bielefeld.de`

Abstract. The purpose of this Wizard-of-Oz study was to explore the intuitive verbal and non-verbal goal-directed behavior of naïve participants in an intelligent robotics apartment. Participants had to complete seven mundane tasks, for instance, they were asked to turn on the light. Participants were explicitly instructed to consider nonstandard ways of completing the respective tasks. A multi-method approach revealed that most participants favored speech and interfaces like switches and screens to communicate with the intelligent robotics apartment. However, they required instructions to use the interfaces in order to perceive them as competent targets for human-machine interaction. Hence, first important steps were taken to investigate how to design an intelligent robotics apartment in a user-centered and user-friendly manner.

Keywords: Social robot ● smart home ● human-robot interaction ● use-case scenario ● usability ● intuitive design ● user-centered design.

1 Introduction

According to Isaac Asimov, "Today's science fiction is tomorrow's science fact." [1]. In order to transform science fiction into science fact, we have to shed more light on determinants of positive user experience and successful interactions with novel technological systems. In the current study, we therefore explored naïve users' interactions with and within a smart home environment that included an assistive robot.

Smart homes are living environments equipped with information technology to assist users in mundane tasks. The smart home stores information about the occupants' needs and habits and utilizes this information to improve the users' comfort, security, and entertainment by connecting the smart home's technology to the world beyond [2,3]. Different smart home solutions with a diverse range of sensors, actuators, and biomedical monitors are already on the market. Individual components can even be

controlled via smartphone or computer [2,3]. For instance, users can check if the oven is turned off when they are not at home. Clearly, a smart environment can be advantageous in many ways. For instance, by monitoring elderly and people with disabilities and by providing assistance to them [3]. Thus, much research has focused on the interaction between these users and ambient intelligence including robots [4]. However, to date, usability research has rather focused on people's interaction with single systems [5]. Other studies have researched technical parameters of smart environments, e.g., sensor data [6] and activity recognition [7]. It is of major importance to provide interfaces according to people's habits and intentions [3]. This enhances the usability and in return the acceptance of such technical devices [8]. Therefore, we have to further explore how people intuitively behave in an intelligent environment.

In the present study, we investigated how naïve users address the cognitive service robotics apartment (CSRA) located at the Cluster of Excellence Cognitive Interaction Technology (CITEC) at Bielefeld University. The smart apartment consists of a kitchen, a living-room, a private gym, and a bathroom. It is complemented by a Meka robot, a bi-manual mobile robot [9] that provides additional assistance. The apartment is equipped with video and audio recording. Additionally, it contains a large amount of sensors to record the user's interaction with the apartment's components and the robot.

1.1   Research Aims

To turn science fiction into science fact, we need to create intuitive, acceptable, and efficient technical systems [2], [8]. To do so, we have to explore how people actually interact with innovation technology upon a first encounter. In the present study, we therefore aimed to shed light on the following research aims: To identify which interfaces people prefer or intuitively address when attempting to complete a given task and to assess how people evaluate the interaction with the robot and the apartment. This leads to related issues, e.g., whether people perceive the apartment and the robot as autonomous entities or whether the system is addressed by name, indicating anthropomorphization. We also wanted to find out more about users' preferences regarding control of the smart environment. Language and gestures are important in interpersonal interaction; thus, verbal and nonverbal communication have an impact on human-machine interaction [10]. Accordingly, the last two research aims considered verbal interaction between the user and the intelligent environment. We investigated if people continue to verbally address their environment after having done so previously. Moreover, we aimed to study if people likewise would address the robot verbally in case it had responded to them in this way. To explore these issues, we conducted an empirical field study to gather a wide range of both qualitative and quantitative data.

2 Method

2.1 Participants

63 participants from Bielefeld University took part in this study. 16 had to be excluded from data analysis due to technical problems or because they did not understand the instructions. All remaining participants ($n = 47$; 25 women, 22 men; $M_{age} = 25.26$, $SD_{age} = 5.69$, age range: 18-50 years) were unfamiliar with the CSRA.

2.2 Procedure

To explore users' intuitive and unbiased behavior in a realistic setup, participants were invited to the CSRA. First, participants gave consent to have their personalized audio and video data recorded. Upon entering the CSRA, the experimenter briefly introduced the apartment, the robot, and the person who had to stay in the apartment during the study for safety reasons (safety person). Participants received cards describing seven mundane tasks which had to be solved in a specific order and as intuitively as possible within the home context. Hence, presumably familiar tasks were chosen which could be solved with each of the given devices. For instance, they had to turn on the light in the hallway (see Table 1 for a full list of tasks). No further information was given about the existing interfaces of the apartment and how they could be operated. To encourage participants' interaction with the intelligent environment, they were told not to use light switches. To reinforce this, all light switches were shut off and no radio or amplifier was available. Furthermore, participants were told not to use their own mobile phones or watches. Additionally, the final task had to be solved without using speech. According to the Wizard-of-Oz setup [11], an experimenter observed participants' attempts from a control room next to the apartment. When detecting a goal-directed action, the experimenter triggered the responses of the system as if the environment was executing the commands autonomously. In advance, only obvious goal-directed actions intending to solve the given task were defined as valid attempts, e.g., a gesture toward the light was interpreted as a signal to switch it on or off. After completing the tasks, participants were asked to complete a questionnaire which took about 10 minutes. Finally, participants were debriefed and reimbursed with €6 or credits for participation. Furthermore, they had the opportunity to ask questions about the study.

Table 1. Seven mundane tasks to be completed in a fixed order.

No.	Order 1	Order 2
1	Turn on the light in the hallway	Turn on the light in the hallway
2	Turn off the light again	Turn off the light again
3	Listen to music	Listen to music
4	Find out if mail has been delivered	Find out if there was a phone call
5	Find out if there was a phone call	Find out the current time
6	Find out the current time	Find out if mail has been delivered
7	Alter the brightness of a light	Alter the brightness of a floor lamp

2.3 Materials

Tasks. Since it was very likely that participants would use spoken language to ask for the mail delivery, the order of tasks was counterbalanced and randomly assigned to participants (Table 1). Moreover, in half of the trials, the robot, the apartment and its features responded using speech when addressed via spoken language. When responding without using speech, the robot and the apartment referred to screens providing text-based information (e.g., indicating the current time).

Questionnaire Data. The questionnaire captured participants' experiences during the interaction with the intelligent environment. Two forced-choice items served to assess which interface participants used most frequently and with which interface they communicated most favorably. Therefore, participants indicated either: 'the robot', 'the apartment', 'both equally', 'none of the two', or 'I don't know'. Moreover, we explored the overall evaluation of the robot and the apartment: how pleasant participants felt during the interaction with the apartment/the robot. We inquired whether participants perceived the robot/the apartment as an autonomous entity. Furthermore, we examined whether participants would have liked to call the robot/the apartment by name. Finally, participants had to indicate whether they had experienced difficulties to solve the given tasks and whether they had felt observed by the safety person. Participants responded to the latter items using a 7-point Likert scale, with high ratings reflecting high endorsement of the measured construct.

Video and Audio Data Analysis. 31 videos had been recorded from three different angles that covered each interactive location of the apartment. In the remaining recordings, a fourth camera was used to provide an additional overview (Figure 1). Besides video and audio material from the kitchen, hallway, and living room, the whole system data were available in separate channels with timing information for each event. System data were temporally aligned with the videos and accessible as annotations [12]. For instance, the Wizard's actions were recorded to identify what the experimenter considered a suitable task solution. More importantly, besides automatically gathered data, the video and audio material was annotated manually to classify participants' behavior during the interaction with the CSRA. In a further step, these annotations allowed statistical data analysis. Annotations were done with EUDICO Linguistic Annotator (ELAN) [13,14]. Recordings were annotated by two raters who created the classification system for participants' behavior. To validate the classification system and to check for consistency between raters, eight videos were annotated by both raters. Inter-rater reliabilities reveal high agreements. Annotation tiers and the respective inter-rater reliabilities are listed in Table 2. Figure 1 gives an overview of the experimental setup and how the annotations according to Table 2 were realized in the ELAN annotation program.

Fig. 1. Four camera perspectives showing the experimental setup including the robot (above left and bottom right camera perspective).

Table 2. Audio and video annotations that depict and classify participants' behavior with inter-rater reliabilities (Cohen's kappa) according to each annotation tier.

Annotation tier	Description
Course of study	Description of single sections of the study depending on the participants' progress and behavior, e.g., if a task was solved successfully or not ($\kappa = 1.00$, $p < .001$).
Method	Participants' method to approach a task e.g., speech, gesture, conventional approach (e.g., using switches), or a combination of multiple methods ($\kappa = 1.00$, $p < .001$).
Focus of attention	Target addressed by the participants before solving a task, e.g., robot, apartment, screens, self-reference, unspecific (unclear addressee) ($\kappa = .69$, $p < .001$).
Final addressee	What participants addressed to solve the task successfully (same options as focus of attention) ($\kappa = .76$, $p < .001$).
Language – address	Description, if participants gave a name to address a target or not ($\kappa = .65$, $p < .001$).
Language – politeness	Indication, if participants addressed a target politely or neutrally ($\kappa = .66$, $p < .001$).
Language – structure	Indication, if participants used concrete questions, phrases or single words ($\kappa = .79$, $p < .001$).
Language – intention	Participants' intention to address an interface, e.g., to greet or to interact with a target ($\kappa = .87$, $p < .001$).
Emotional expression	Type of emotion expressed by participants (only if an emotion was particularly apparent, e.g., happiness, fear. Therefore, inter-rater reliability could not be considered).

3   Results

ELAN-annotations documented and classified participants' behavior during the interaction with the intelligent environment. Based on these annotations, we focused on absolute and relative frequencies to analyze which interface participants addressed and which approach they used to complete each task. Questionnaire and video data were analyzed by computing t-tests, Chi-square, and absolute and relative frequencies (%). To compare participants' behavior during the study to their responses to the questionnaire, both results will be reported to establish convergent validity.

3.1 Participants' Addressees and Methods

Table 3. Interfaces used per task.

Task	Interface	Absolute frequency	Relative frequency
Switch on the light	light in the hallway	26	55.3
	screen	8	17
	unspecific	7	14.9
	robot	5	10.6
	switch	1	2.1
Switch off the light	light in the hallway	25	53.2
	sliding-door	6	12.8
	robot	5	10.6
	unspecific	4	8.5
	screen	3	6.4
	general switch	2	4.3
	self-reference	1	2.1
	apartment	1	2.1
Play music	unspecific	24	51.1
	robot	6	12.8
	screen	6	12.8
	speaker	5	10.6
	not solved	1	2.1
	self-reference	1	2.1
	general switch	1	2.1
	unclear, if addressed sth	1	2.1
	apartment	1	2.1
	electronic switch	1	2.1
Ask for a delivery	unspecific	17	36.2
	robot	14	29.8
	screen	12	25.5
	self-reference	2	4.3
	fitment	2	4.3
Ask for a phone call	unspecific	25	53.2
	robot	14	29.8
	screen	5	10.6
	not solved	2	4.3
	self-reference	1	2.1
Ask for the current time	unspecific	16	34
	screen	14	29.8
	robot	13	27.7
	self-reference	2	4.3
	fitment	1	2.1
	apartment	1	2.1
Alter brightness of a light	floor lamp	41	87.2
	screen	4	8.5
	robot	1	2.1
	general switch	1	2.1

Table 3 focusses on the first research aim regarding which interfaces participants would intuitively and most frequently address. Whenever the task referred to a physical interface (e.g., control the light in the hall way), most participants addressed this entity directly. Compared to the apartment, the robot was addressed more frequently regardless of the task, but the addressee often remained unspecific. That means it was obvious that participants addressed an interface within the apartment, but it was unclear which one. To investigate this finding, we considered additional questionnaire data. According to

participants' statements, absolute and relative frequencies reveal that they addressed the apartment most frequently (χ^2 (3, $N = 47$) = 14.00, $p = .003$) and by tendency most favorably (χ^2 (4, $N = 47$) = 7.79, $p = .10$) compared to the robot (Table 4). With regard to the second research aim, it turned out that people assessed the interaction with the apartment as more pleasant compared to the interaction with the robot ($M_{apartment} = 5.61$, $SD_{apartment} = 1.56$, $M_{robot} = 4.48$, $SD_{robot} = 1.75$, $t(45) = 4.92$, $p < .001$, $d = 0.68$).

Table 4. Comparison between the interaction with the robot and the apartment.

	Most frequently addressed		**Most favorably addressed**	
	Absolute frequency	Relative frequency	Absolute frequency	Relative frequency
Apartment	17	36.2	21	45.7
Robot	7	14.9	8	17.4
Both equally	7	14.9	4	8.7
None	8	17	13	28.3
Do not know	8	17	0	0

Since the apartment and the robot were programmed to reply according to a given task (e.g., to say or to indicate that there was a phone call), we investigated if participants perceived them as an autonomous entity. The survey data revealed that participants neither perceived the robot, nor the apartment as an autonomous entity ($M_{apartment} = 2.66$, $SD_{apartment} = 1.79$, $M_{robot} = 2.85$, $SD_{robot} = 1.60$, $t(46) = -.62$, $p = .535$, $d = 0.11$). This leads to the investigation, whether participants would have liked to address the robot/the apartment by name. In the initial instructions, neither the apartment, nor the robot had been introduced to the participants. Survey responses showed that participants would have rather addressed the robot by name than the apartment ($M_{robot} = 5.09$, $SD_{robot} = 2.00$, $M_{apartment} = 3.40$, $SD_{apartment} = 2.37$, $t(46) = -5.50$, $p < .001$, $d = 0.71$). Comparing this finding to the video data, we found that only three out of 47 participants named an entity to address it, e.g., "Could you help me, robot?" In these interactions, the robot was four times addressed particularly, the hallway light and the apartment were each addressed twice.

To explore which method participants used to solve a given task, absolute and relative frequencies were computed (Table 5). Each task was most frequently solved by verbal interaction with the environment, except when participants were explicitly told not to use speech to alter the brightness of the floor lamp. In the latter case, most participants used gestures or touched the lamp. 41 of the 47 participants used speech to control their environment, regardless of the task order and the environmental feedback (verbal vs. non-verbal). Moreover, all those who used speech once, continued to interact verbally with their environment. These findings are conform to participants' statements from the questionnaires. When asked which interface they would mainly use in an intelligent apartment, most participants stated they would use speech (28 participants, 73.3%) followed by tablet/laptop interfaces (8 participants, 21.1%), and the robot (2 participants, 5.3%). None of the participants chose to mainly use gestures and facial expressions to control the intelligent environment.

Table 5. Absolute and relative frequencies of task completion approaches.

Task	Method	Absolute frequencies	Relative frequencies
Turn on the light	Speech	26	55.3
	Gesture	10	21.3
	Touch	8	17
	Locomotion	1	2.1
	Search behavior	1	2.1
	Combination of methods	1	2.1
Turn off the light	Speech	31	66
	Touch	7	14.9
	Gesture	6	12.8
	Combination of methods	2	4.3
	Search behavior	1	2.1
Play music	Speech	38	80.9
	Touch	6	12.8
	Gesture	2	4.3
	Not solved	1	2.1
Ask for a delivery	Speech	40	87
	Touch	2	4.3
	Conservative methods	2	4.3
	Gesture	1	2.2
	Search behavior	1	2.2
Ask for a phone call	Speech	40	85.1
	Touch	3	6.4
	Not solved	2	4.3
	Gesture	1	2.1
	Search behavior	1	2.1
Ask for the current time	Speech	40	87
	Gesture	3	6.5
	Touch	2	4.3
	Search behavior	1	2.2
Alter the floor lamp	Touch	22	46.8
	Gesture	21	44.7
	Speech	2	4.3
	Conservative methods	1	2.1
	Combination of methods	1	2.1

According to our last research aim, most of those who had interacted with the robot once, continued to interact with it instead of trying another interface. Remarkably, only 15 (10 women, 5 men) participants used the robot to solve a given task. Although all of those had successfully completed the task, two did not continue their interaction with it. Finally, participants indicated that they had no difficulty to solve the given tasks ($M = 2.74$, $SD = 1.34$), but felt markedly observed by the safety person ($M = 4.49$, $SD = 1.98$).

4　Discussion

The current research highlights the importance of user-centered studies in the development process of smart homes. Naïve users had to fulfill seven mundane tasks in an intelligent apartment. Participants were instructed to behave intuitively and without using conventional approaches (e.g., light switches). They were not explicitly introduced

to the interfaces of this environment. Audio and video data recording participants' interaction were supplemented by qualitative questionnaire data which assessed participants' evaluations of the interaction with the environment. Recordings showed that regardless of the task order or whether the system gave verbal or non-verbal feedback, participants preferred physical interfaces whenever the task allowed to (e.g., to control the light in the hallway). Only a minority of participants addressed the robot. Although regardless of the task, the robot was addressed more frequently than the apartment. However, it was often unclear which interface was addressed. Questionnaire data might shed light on these findings. In these, participants stated to have addressed the apartment more frequently and preferentially than the robot. Maybe they were not aware to have interacted with the apartment as an autonomous interface, until the questionnaire pointed to it as such. In line with this, only two participants referred to the apartment particularly. Additionally, the questionnaires revealed that participants would have preferred to call the robot by name compared to the apartment. Interestingly, the interaction with the apartment was perceived as more pleasant compared to the robot. Thus, participants might have addressed both devices more frequently and namely, if they would have been introduced in particular. Regardless of the task order or feedback, most participants used speech to control the smart home environment. Except when they were explicitly told not to use speech to alter the brightness of the floor lamp, most participants used gestures or touches. Those who verbally interacted with the environment once, continued to do so. Similarly, most participants who interacted with the robot once, continued to address it. Some addressed it just to try if it responds. Since it only responded to an appropriate task solution, participants might not have considered it as a competent addressee as it did not signal attention to the participant [15]. After the study, many participants voiced regret that they had not interacted with the robot or the apartment, whereas during the study, this did not seem like an option to them. Others explained they were afraid of the robot because of the presence of a safety person. Moreover, they were afraid the robot could move towards them unexpectedly. Participants did not indicate difficulties to solve the given tasks, but some stated they did not dare to try some methods because they felt observed by the safety person.

Summing up, this research makes an important contribution to the existing literature on ambient intelligence by providing empirical evidence based on a multi-method approach to validate the developments within CSRA by means of a user-centered approach [8]. The present study shows that naïve users are able to find innovative and reasonable methods to interact with the intelligent apartment. Nonetheless, information regarding the smart home's interfaces should be provided in order to perceive the apartment and its equipment as 'smart' interfaces. This might help to overcome people's safety concerns, particularly when interacting with the robot. Hence, personalizing the robot through introducing it namely could remedy this issue. Further, providing information about the environments' interfaces might enhance people's trust in this technology and therefore enhance their motivation to interact with it [8]. The current findings make clear that to enable a smooth communication with smart environments, interaction patterns should resemble interpersonal interaction (e.g., including speech and gestures). At the same time, conventional interfaces should remain available. Thus, we recommend to combine intuitively controllable interfaces with conventional and directly addressable ones. Further work is needed to identify the most optimal way of

introducing and designing the various interfaces. For the time being, this research provides important answers how to design intelligent robotics apartment. Therefore, it makes an important contribution to transform science fiction into a science fact.

References

1. Asimov, I.: AZ Quotes, `http://www.azquotes.com/quote/877722`
2. Kaasinen, E., Kymäläinen, T., Niemelä, M., Olsson, T., Kanerva, M., Ikonen, V.: A User-Centric View of Intelligent Environments: User Expectations, User Experience and User Role in Building Intelligent Environments. Computers 2, 1-33 (2012)
3. Chan, M., Campo, E., Estève, D., Fourniols, J.Y.: Smart homes - current features and future perspectives. Maturitas 64, 90-97 (2009)
4. Pineau, J., Montemerlo, M., Pollack, M., Roy, N., Thrun, S.: Towards Robotic Assistants in Nursing Homes: Challenges and Results. Robot. Autonom. Syst. 42, 271-281 (2003)
5. Venkatesh, V., Davis, F.D: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Manage Sci. 46, 186-204 (2000)
6. Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., Devlin, S.: Evidential Fusion of Sensor Data for Activity Recognition in Smart Homes. Pervasiv. Mobile Comput. 5, 236-252 (2009)
7. Chen, L., Nugent, C.D., Wang, H.: A Knowledge-Driven Approach to Activity Recognition in Smart Homes. IEEE Transactions on Knowledge and Data Engineering, pp. 961-974. IEEE Press, New York (2012)
8. Pavlou, P.A.: Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. Int. J. Electron. Comm. 7, 101-134 (2003)
9. MeKa Robotics: Aaron Edsinger, Jeff Weger, San Francisco (2006)
10. Moore, R.K.: PRESENCE: A Human-Inspired Architecture for Speech-Based Human-Machine Interaction. IEEE Transactions on Computers, pp. 1176-1188. IEEE Press, New York (2007)
11. Kelley, J.F.: An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. T Inform. Syst., 2, 26-41 (1984)
12. Holthaus, P., Leichsenring, C., Bernotat, J., Richter, V., Pohling, M., Carlmeyer, B., … Wrede, S.: How to Address Smart Homes with a Social Robot? A Multi-Modal Corpus of User Interactions with an Intelligent Environment. In: 10th Edition of the Language Resources and Evaluation Conference, LREC Press, Portoroz (2016)
13. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands: ELAN, `http://tla.mpi.nl/tools/tla-tools/elan/`
14. Sloetjes, H., Wittenburg, P.: Annotation by Category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC Press, Marrakech (2008)
15. Holthaus, P., Pitsch, K., Wachsmuth, S.: How Can I Help? Int. J. Soc. Robot. 3, 383-393 (2011)