
Why UX Research Matters for HRI: The Case of Tablets as Mediators

Jan de Wit

Tilburg University
Tilburg, the Netherlands
j.m.s.dewit@uvt.nl

Rianne van den Berghe

Utrecht University
Utrecht, the Netherlands
m.a.j.vandenberghe@uu.nl

Paul Vogt

Tilburg University
Tilburg, the Netherlands
p.a.vogt@uvt.nl

Laura Pijpers

Tilburg University
Tilburg, the Netherlands
laura.pijpers93@gmail.com

Emiel Krahmer

Tilburg University
Tilburg, the Netherlands
e.j.krahmer@uvt.nl

ABSTRACT

Many human-robot interaction systems involve a third component: a tablet, which can either be separate or integrated in the robot (as is the case in SoftBank Robotics' Pepper robot). Such a tablet can be used, for instance, to present information to the human user or to gain control over the robot's complex surroundings, by introducing a virtual environment as a substitute for interactions that would normally happen in the physical world. While such a tablet can potentially have a big impact on the usability of the entire system and affect the interaction between human and robot, it is often not explicitly included when evaluating the user experience of human-robot interaction. This paper

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI2019 SIRCHI Workshop, May 2019, Glasgow, UK

© 2019 Copyright held by the owner/author(s).

describes a case study where three evaluation methods were combined in order to get a comprehensive overview of the user experience of an Intelligent Tutoring System (ITS), consisting of a robot and a tablet. The results show several major usability issues with the virtual environment, which could have affected the experience of interacting with the robot. This underlines the importance of including not only the robot itself, but also any other interaction mediators in an iterative design process.

CCS CONCEPTS

• **Human-centered computing** → Usability testing; Heuristic evaluations; Interaction design process and methods; • **Applied computing** → *Interactive learning environments*; • **Computer systems organization** → *Robotics; Robotic autonomy*.

KEYWORDS

Autonomous robots; user centered design; design methodology; human-robot interaction

ACM Reference Format:

Jan de Wit, Laura Pijpers, Rianne van den Berghe, Emiel Kraemer, and Paul Vogt. 2019. Why UX Research Matters for HRI: The Case of Tablets as Mediators. In *Proceedings of the Workshop on the Challenges of Working on Social Robots that Collaborate with People, ACM CHI Conference on Human Factors in Computing Systems (CHI2019 SIRCHI Workshop)*. ACM, New York, NY, USA, 6 pages.

INTRODUCTION

Robots are increasingly being used for application domains in which they are expected to interact frequently with humans, and thus to exhibit socially intelligent behavior. Examples of such domains include personal assistance, education, and health care [4]. Socially intelligent behavior relates to aspects such as expressing and perceiving emotions, communicating with high-level dialogue, establishing and maintaining social relationships, using natural cues such as gaze and gestures, showing personality and character and displaying the ability to learn or develop social competencies [5]. In order to be perceived as being social, a robot will need to exhibit at least a certain degree of autonomous behavior [2], which consists of its ability to sense, plan and act in its environment, with the intent of reaching a task-specific goal without external control [3]. Social robots that are able to operate fully autonomously can be deployed in a range of real world settings, and contribute to research in human-robot interaction (HRI) by allowing studies to be conducted in the field – at home, school, health care facilities – over longer periods of time without having a researcher control the robot’s behavior. This in turn leads to higher ecological validity, reduces bias and improves the replicability of these studies.

However, autonomous social behavior is challenging to implement. The sensing abilities of social robots include the observation of not only the robot’s often complex and unpredictable physical

surroundings, but also the characteristics and behavior of humans that are present in this environment. Human social behavior is a complex phenomenon that is still under research, which adds to the challenge of developing a robot that is able to interact socially with humans [9]. It is unclear exactly which types of sensing, planning and acting functionalities are needed to facilitate social interactions. Moreover, some of the techniques that are currently being used do not perform well enough to be used autonomously, in a complex environment. Finally, successful completion of the robot's task-specific goals can be difficult to measure, when these goals involve a change in knowledge state, attitude or behavior of a conversational partner.

One way to cope with these challenges is to control or constrain the environment, thus making it easier for the robot to sense and act within its surroundings. This can be done by moving part of the human-robot interaction into the virtual domain, for example by introducing a tablet device as a mediator. Objects within the virtual space can easily be tracked and manipulated programmatically by robots, as well as through a graphical user interface by humans, thereby allowing both parties to collaborate on the device in order to complete their tasks. However, what is often not critically evaluated is how the introduction of such a virtual environment may influence the overall user experience of the human-robot interaction, and to what extent it diminishes the benefits of the robot's physical presence in a real world context. If such peripherals are being used, they should also be involved in an iterative design process, included when setting user experience goals, and subject to user experience evaluation methods.

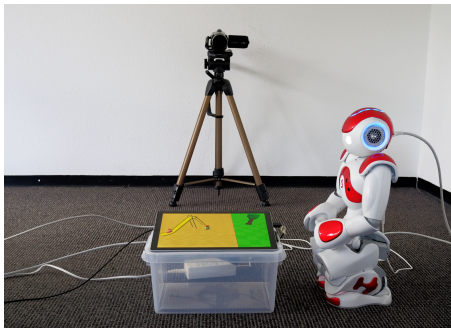


Figure 1: The setup of the Intelligent Tutoring System (published with permission from [7]).



Figure 2: The virtual environment used in lesson one of the intelligent tutoring system.

OUR CASE STUDY

We have recently conducted a longitudinal tutoring study where pre-school children (five to six years old) learned second language vocabulary by participating in six lessons during which they were performing a number of interactions with the robot and a tablet device. Figure 1 shows the experimental setup, Figure 2 is an example of a scene that was displayed on the tablet device during a lesson, and Figure 3 illustrates the number of different interactions per lesson, which included various manipulations of objects in the virtual environment, repeating after the robot and physically enacting certain concepts. The study included a tablet condition, where children interacted only with the tablet device and the robot's speech output was routed through the tablet's speakers. Although existing research suggests a beneficial effect of the robot's physical presence on learning, we found no significant difference between the tablet condition and the conditions where the robot was physically present [7].

To learn more about the interactions with the system and how this may have affected our findings, we conducted a usability and user experience evaluation of the Intelligent Tutoring System (ITS) as a whole. Due to time constraints, we limited our scope to the first and last lesson in the series, allowing us to take into account the learnability of the system. A triangulation approach [8] was used

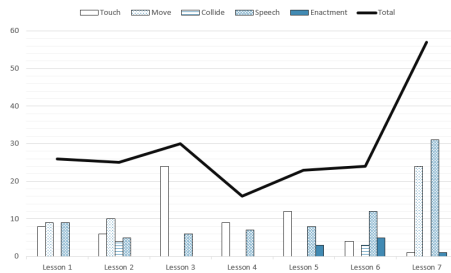


Figure 3: The number of interactions per lesson, split by interaction type.

Table 1: Performance measures extracted from log files. The number of errors is divided by the number of interactions of that type that were present in the lesson.

	Lesson 1	Lesson 6
Time on task (touch)	22.57 sec	7.70 sec
Time on task (move)	28.99 sec	9.64 sec
Errors (touch)	5	2.75
Errors (move)	50.89	N/A
Task success (touch)	97.9%	98.4%
Task success (move)	68.3%	96.7%

to combine the outcomes of three different evaluation methods. First, observations were conducted on a set of recordings of child-robot interactions from the aforementioned longitudinal study [7], and performance metrics such as time on task were extracted from the log files belonging to these interactions. A random selection was made of 60 out of the 162 children that participated in the experimental conditions. Second, three design experts were asked to evaluate the system based on the Heuristic Evaluation Child E-learning applications (HECE) [1]. Finally, the ITS was evaluated with ten older children (11–12 years old) by means of a think aloud session with the system, followed by a semi-structured interview. For efficiency reasons, eight children were invited in groups of two and the other two individually, ensuring that lessons one and six would each be rated five times. The combination of these three methods resulted in quantitative measurements of children’s performance, qualitative feedback on the user experience and a list of usability issues. The Damage Index [6] was calculated in order to consolidate the reported severity ratings of each usability issue from multiple methods. This measure takes into account the average severity rating across methods, as well as the number of methods in which the issue came up. As a result, issues that occur frequently are assigned a relatively high Damage Index compared to issues that are more rare.

RESULTS

Observations and Log Files. The observations resulted in a list of usability issues for lessons one and six, along with the number of times in total, and for how many individual children, each issue occurred. Two researchers assigned a severity rating to each issue. A total of 44 out of the 80 issues (55%) had a severity rating of 1 (prevents task completion) or 2 (significant delays and frustration). All of these were related to tasks where the child had to move an object to a new location, or to collide it with another object. The performance measures, which were extracted from the log files collected by the system, are shown in Table 1.

Heuristic Evaluation. The heuristic evaluation resulted in a list of 25 issues, of which eleven (44%) received the highest two severity ratings. These issues were related to tasks that could not be properly carried out on the tablet (due to bugs), a lack of feedback for tasks where the child has to enact something (e.g., raising their right hand), unclarity in the robot’s pronunciation of certain words, the imposed, slow pace of the interaction, the design choices regarding the tablet game (2D was suggested over 3D) and the lack of introduction of certain game mechanics (the screen turning black to guide the child’s attention).

Usability Study and Interview. Issues that the children encountered while interacting with the system and suggestions that they made were noted down, and a severity rating was assigned to these points by two of the researchers. From a total number of forty issues, five were assigned a severity rating of 1 or 2 (12.5%). These were related to problems when moving objects, difficulty understanding the

Table 2: Number of usability issues per Damage Index range.

< .11	.11-.2	.21-.3	.31-.4	.41+
5	14	3	7	6

¹<https://bit.ly/2E9UTK4>

robot's speech and confusion about what kind of action was expected of the user. The other comments overlap to a large extent with the findings from the heuristic evaluation (e.g., unclarity of gestures, and overall pacing of the interaction).

Combined results. The lists of usability issues resulting from the three different evaluation methods were combined, where overlapping issues were merged. This resulted in a total of 35 unique issues. To get an overview of the severity of each issue, taking into account the number of evaluation methods in which it was reported, and the severity that was assigned in these methods, a Damage Index was calculated. Table 2 shows the number of issues belonging to different Damage Index ranges. The top thirteen issues with a Damage Index of at least 0.31 were related to problems with dragging objects on the tablet, tasks not being clear to the user, the slow and fixed pacing of the interaction, limited control of the user over the system, unnatural and unclear speech from the agent, interaction mechanics not being properly introduced, ambiguous words or gestures, lack of feedback from the agent, and objects on the screen being locked while the agent is talking. The full results are made available online¹.

DISCUSSION

As we work towards creating social robots that are capable of operating fully autonomously in a complex and dynamic environment, we attempt to exercise control over this environment in order to deal with technical limitations and the intricacies of social interactions. The results of the current evaluation show that the use of a virtual environment as a mediator for human-robot interactions can greatly affect the overall user experience. Most issues reported were either directly related to the interactions with the tablet (such as issues with moving objects on the screen), or listed as issues with the robot although they can actually be traced back to the tablet, because the robot was provided with incorrect information regarding the state of the virtual environment. This in turn resulted in the robot performing incorrect actions based on erroneous input, such as saying the wrong things or giving negative feedback when the task was in fact completed successfully.

A triangulation approach combining three different methods helped to get a comprehensive overview of issues that occurred with the system as a whole. Out of the 35 issues in the combined list, 22 were identified in only one of three evaluation methods. This is an indication that a substantial amount of issues would not have been identified if we had used only one evaluation method. However, in this case the evaluations were conducted *after* the system was already used in a large-scale experiment [7]. It is now impossible to tell whether, and to what extent, the findings from that experiment were influenced by the issues encountered when evaluating the system, without running a similar experiment after resolving these issues. In future work, we would therefore start evaluating earlier in, and more frequently throughout the design process.

CONCLUSION

This paper describes a case study in which a usability and user experience evaluation of an Intelligent Tutoring System was conducted. The study underlines the importance of evaluating the overall experience of a human-robot interaction, including any mediating devices that are introduced to gain control over the robot's environment in order to increase the robot's level of social autonomy. Furthermore, we urge researchers to allocate resources to the design and development of such interaction mediators, and to report exactly to which degree their robot is able to behave autonomously, as well as any concessions or work-arounds that might be in place. This would ensure that the effects of any mediators on experimental findings are minimized, while at the same time providing the research community with enough information to be able to reproduce these findings.

ACKNOWLEDGMENTS

This research is conducted as part of the L2TOR project, which has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 688014.

REFERENCES

- [1] Asmaa Alsumait and Asma Al-Osaimi. 2009. Usability heuristics evaluation for child e-learning applications. In *Proceedings of the 11th international conference on information integration and web-based applications & services*. ACM, 425–430.
- [2] Christoph Bartneck and Jodi Forlizzi. 2004. A design-centred framework for social human-robot interaction. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication, RO-MAN*. IEEE, 591–594.
- [3] Jenay M Beer, Arthur D Fisk, and Wendy A Rogers. 2014. Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction* 3, 2 (2014), 74–99.
- [4] Kerstin Dautenhahn. 2007. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362, 1480 (2007), 679–704.
- [5] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and autonomous systems* 42, 3-4 (2003), 143–166.
- [6] Gavin Sim and Janet C Read. 2010. The damage index: an aggregation tool for usability problem prioritisation. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*. British Computer Society, 54–61.
- [7] Paul Vogt, Rianne van den Berghe, Mirjam de Haas, Laura Hoffmann, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, Ora Oudgenoeg-Paz, Fotios Papadopoulos, Thorsten Schodde, Josje Verhagen, Christopher D. Wallbridge, Bram Willemsen, Jan de Wit, Tony Belpaeme, Kirsten Bergmann, Tilbe Gökşun, Stefan Kopp, Emiel Kraemer, Aylin C. Küntay, Paul Leseman, and Amit K. Pandey. 2019. Second Language Tutoring using Social Robots: A Large-Scale Study. In *the 2019 ACM/IEEE International Conference on Human-Robot Interaction*.
- [8] Chauncey E Wilson. 2006. Triangulation: the explicit use of multiple methods, measures, and approaches for determining core issues in product development. *Interactions* 13, 6 (2006), 46–ff.
- [9] Guang-Zhong Yang, Jim Bellingham, Pierre E Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, Vijay Kumar, Marcia McNutt, Robert Merrifield, et al. 2018. The grand challenges of Science Robotics. *Science Robotics* 3, 14 (2018), eaar7650.