

User Experience Evaluation with a Wizard of Oz Approach: Technical and Methodological Considerations

A. Weiss*, R. Bernhaupt**, D. Schwaiger*, M. Altmaninger*, R. Buchner*, M. Tscheligi*

Abstract—User experience evaluation in human-robot interaction is most often an expensive and difficult task. To allow the evaluation of various factors and aspects of user experience, a fully functional (humanoid) robot is recommended. This work presents technical and methodological considerations on the applicability of the Wizard of Oz (WOz) approach to enable user experience evaluation in the field of Human-Robot Interaction. We briefly describe the technical aspects of the set-up, the applicability of the method, and a first case study using this methodological approach to gain an early understanding of the user experience factors that are important for the development of a human-humanoid interaction scenario.

I. INTRODUCTION

“Imagine you are working at a construction site and you receive the task from your principal constructor to mount a gypsum plasterboard in collaboration with a humanoid robot. You can control the robot with predefined voice commands”. The evaluation of user experience (UX) factors in human-robot collaboration is a difficult task during the early development stages. User experience is still a loosely defined term in human-computer interaction, but in general it refers to all experiences a user has before, during, and after interacting with an (interactive) product [8]. The term user experience must not be confused with usability. User experience goes beyond efficiency, effectiveness, and satisfaction that is felt when interacting with a system [14], but refers to concepts like emotion, affect, fun, enjoyment, beauty, and other hedonic attributes [13]. To understand the users’ experiences when interacting with the robot, a variety of methods is used.

To allow a realistic impression of the interaction with a system or robot, user experience evaluation is most often conducted with fully functional (prototypical) systems using questionnaires to evaluate the users’ experiences after interacting with the system. For the above mentioned construction site scenario we would need a fully functional robot to evaluate user experience in a realistic setting. This approach is expensive and only allows evaluation at late development stages. Additionally (provided that a fully functional robot is available), the evaluation of a collaborative task in a real construction site might not be possible due to security issues. To close this methodological gap of user experience evaluation for early development stages, we propose the usage of a Wizard of Oz approach.

Evaluation of user experience of (new forms of) interaction techniques in human-robot interaction is affected by various factors. To allow the evaluation of user experience, we have to consider that the development of robots is typically not iterative and based on user-centered design, but robot development is more often use-centered [5]. User experience evaluation methods from traditional Human-Computer Interaction (HCI) thus might not be applicable and useful for the development of robots. User experience factors should be evaluated during the design phase of the robot to allow a successful implementation of aspects supports an overall positive user experience.

Looking at the findings on multimodal interaction in the field of HCI, it still remains unclear, to which extent we can use these findings on overall user experiences when looking at users interacting with a robot. Contrary to standard interactive systems (with typically a screen allowing to interact and receive feedback), a humanoid robot can be touched by the user and interaction is more human-human like than any other form of cooperation with interactive systems. The ability to touch a robot and the expressions and gestures a robot can show, change the interaction of users. Thus, findings from the area of HCI on user experience aspects of multi-modal interaction might not be transferable to the HRI domain.

To understand how users perceive the interaction and collaboration with a robot in general, we argue that it is necessary to evaluate user experience factors early in the design phase, and therefore propose a Wizard of Oz approach (WOz) as it allows the evaluation of UX at such early phases.

The goal of this work is to describe how to set up a Wizard of Oz approach using mixed-reality which enables user experience evaluation of new forms of multimodal interaction techniques and to show that the WOz approach is realistic enough to evaluate different interaction techniques in terms of UX.

The rest of the paper is structured as follows: First, we discuss related work on user experience evaluation in the field of HCI, and we describe methodological limitations when applied to the field of HRI. Next, we propose the WOz approach as a possible methodological approach to evaluate user experience in HRI at early design phases, presenting a brief technical description of the set-up. Finally, we describe a first evaluation study to show the applicability of the method and summarize (methodological) lessons learned during this case study.

M. Altmaninger, D. Schwaiger, R. Buchner, A. Weiss, and M. Tscheligi are with the HCI and Usability Unit, ICT&S Center, University of Salzburg, 5020 Salzburg, Austria `firstname.lastname@sbg.ac.at`

R. Bernhaupt is with IRIT, Groupe IHCS, 118 Route de Narbonne, 31062 Toulouse Cedex 9 `regina.bernhaupt@irit.fr`

II. RELATED WORK

Human-Computer Interaction offers a broad variety of user experience evaluation methods. User experience evaluation methods range from questionnaires [8] to bio-physiological measurements [15] and aim to evaluate aspects like fun, enjoyment, flow, beauty, hedonic quality, emotions, affects, and moods. Most of the evaluation methods are applied in lab or field studies, allowing the user to interact with a real prototype.

The applicability of these methods for human-robot interaction is limited. Prototyping human-robot collaboration (HRC) with a robot is especially hard if it involves a humanoid robot. Dautenhahn et al. presented a sketch of a typical development timeline of robots intended to collaborate with humans (see [4]). In an initial phase of planning and specification, mock-up models might be used before hardware and software development commences [1].

Wizard of Oz refers to a range of methods in which some or all of the interactivity that would normally be controlled by computer technology is “mimicked” or “wizarded”. It is considered to be a mainstream method in HCI and, as user groups have diversified and the technologies under investigation have changed, the Wizard of Oz method has become a feature of many studies. In a traditional Wizard of Oz study, there is a human wizard who manipulates the interface or “wizards” the interaction technique in the human-robot interaction. In WOz studies in Human-Robot Interaction research the response behaviour of embodied robots is often replaced by a wizard approach (see eg. [9]).

In Human-Computer Interaction the WOz technique was used in the past to understand new forms of interaction techniques, especially multimodal forms which were too difficult to develop (see e.g. [11]). Since then the WOz technique has been extensively used to validate and investigate (multimodal) interaction techniques including various forms of feedback.

Our work is related to the usage of the WOz technique in augmented reality settings [10], but extends the augmented reality to a mixed-reality setting by allowing the user to physically interact with a simulated humanoid robot when conjointly lifting a board (including force feedback).

We argue that from the experimental perspective the WOz approach proposed in the following allows to simulate the real interaction with a humanoid robot to a reasonable extend and thus enables the evaluation of user experience aspects.

III. USER EXPERIENCE EVALUATION WITH WOZ

The goal of this WOz evaluation set-up is to provide insights into the overall user experience when collaborating with a robot using a multimodal interaction technique that consists of speech and several forms of feedback including force feedback. The basic concept for the WOz approach is task based: A human worker and the humanoid HRP-2 robot collaboratively pick up, move, and mount a board. The robot can be controlled by voice commands and by haptic input (pushing and pulling of the board). The human co-worker receives haptic feedback. In a human-human interaction the

person who currently has the overview of the situation, would navigate the other one by means of voice commands, pushing or pulling the object into the right direction, and gestures to signal obstacles.

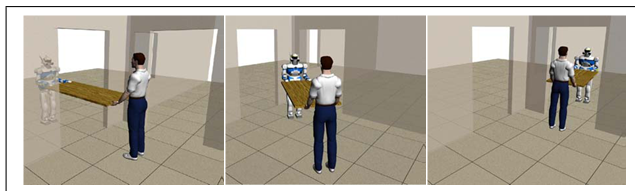


Fig. 1. Human-Robot Collaboration Scenario

Figure 1 shows an already rather complex implementation of this task for human-robot collaboration. In the first step the robot directs the task (robot: leader, human: assistant), whereas at the end the situation changes and the human is the leader (robot: assistant, human: leader). The complex element of this task is that the collaboration between the human and the robot is based on haptic contact via the board and not on direct contact interaction. Thus, the assistant has to follow the directions of the leader that are communicated via the motions of the board and/ or speech commands. Because of the change in the leader and assistant situation, the feedback modalities of the robot are of high relevance.

To allow to understand user experience aspects for this type of interaction technique (speech and haptic feedback), the task was specified as follows: A human user should mount a board together with the 3D model of the humanoid HRP-2 robot.

- 1) The robot needs to be told to move to the spot (in front of the board) where the collaboration starts.
- 2) The board needs to be lifted together.
- 3) The board needs to be moved (by a side step motion) to another place.
- 4) The board needs to be tilt forward to a column together with the robot.
- 5) The robot needs to be told to screw the board.

The main requirement for the simulation was to enable the user to interact with the simulated HRP-2 robot in an intuitive way, additionally supported by different feedback modalities. The prototypical implementation should allow the user to understand how the interaction with a real robot would be. To support a wide variety of interaction possibilities, we decided to prototypically implement four modalities, which can be used to interact and collaborate with the robot:

- direct manipulation of the board using a real gypsum plaster board as input device
- speech recognition of the robot
- visual feedback
- force feedback

In the following we describe how to implement this WOz scenario from a technical perspective, followed by a brief experimental pre-study of the scenario, showing how to use the WOz for UX evaluation.

IV. TECHNICAL IMPLEMENTATION

From the technical implementation side our WOz approach is new in terms of combining direct manipulation including force feedback with a 3D implementation of a humanoid robot based on a game engine. The usage of a game engine for experience prototyping of human-robot collaboration offers several advantages: Common game engines are well supported by their community and offer a wide range of tools, which enables a fast and inexpensive way to create simulated environments. The simulation created for the human-robot collaboration scenario with HRP-2 was realized as a modification of the game Crysis. Crysis delivers a framework with many features including an application programming interface to create customized game elements. For a typical augmented reality WOz study the experimental set-up has to be described, including the methodological set-up of the used instruments for measurement (1). A WOz study additionally needs [10] a tool for capturing the user data (2), a possibility to observe and/ or measure the interaction technique (3), and a support for the remote control of the wizard (4).

A. The Setting

To ensure a “close-to-real-experience-prototype” which enables the evaluation of user experience aspects of the human-robot collaboration scenario, an augmented reality simulation was set up. For this purpose we decided to split the presentation of the scenery in two parts divided by a screen. On one side there was the simulated robot placed in the construction site presented by the game engine. On the other side the user was interacting with the simulated robot via a half “real” half “simulated” plaster board. Other bridging elements between the “real” action space of the human and the “simulated” action space of the robot, were a table where the board was placed in the beginning and a wall where the board had to be mounted at the end. This enabled the users to interact with the simulation and manipulate the 3D simulated part of the test scenario in a direct way (see figure 2).



Fig. 2. Haptic Augmented Simulation Setting

Several modifications to the bone system of the Crysis engine were made to adapt the robot’s movements. The virtual skeleton of the robot was prepared to be connected to several different key points. These points offered the possibility to control the simulated HRP-2 model similar to a string puppet. This technique offered a real time reaction of the robot to the movements that were performed by the test participants. The state of the robot’s bone structure was automatically adapted in real time. Further, the robot “listened” to a set of voice commands. Each voice command triggered a specific predefined action sequence. Thus, the robot was controlled with a semi-automatic approach, ensuring the adaption of the simulation as well as the comparability between the participants’ performances.

1) *Direct Manipulation of the Plaster Board:* To capture each movement of the plaster board, a Wii remote control was strapped onto the board. The sensor data of the remote was used to synchronize the movements of the board outside of the 3D simulation with its virtual extension. In the virtual scene the robot grabbed the board and reacted to every movement of it. This extension of the real board into the screen created the illusion of actually lifting the board in collaboration with the simulated HRP-2 robot.

2) *Speech Recognition:* Instead of using speech recognition we had the wizard to simulate real speech recognition. As the goal of the WOz study was to understand user experience aspects of a final robot (with an excellent speech recognition), we considered the simulation of speech recognition as advantageous. The voice commands (typed in by the wizard) triggered different action states in the simulation. On the contrary, the actions were not controlled by the wizard, but were scripted action sequences, to ensure that the robot reacted consistently on the actions of each participant in the experimental setting.

B. The Feedback Modalities

Glencross et. al [6] argue that a combination of the following four factors is required for credible virtual reality environments.

- 1) high fidelity graphics
- 2) complex interaction engaging multiple sensory modalities
- 3) realistic simulations
- 4) state of the art tracking technology

Thus, developing simulations as applications in virtual reality requires adequate feedback and interactivity. As we simulate aspects of the interaction rather than technical conditions, the complexity of the interaction directly influences the realism of the simulation. To enhance the realism of this sort of simulation, the interaction modalities should support the “close-to-real-experience”. Therefore, a representative feedback system is the key factor to achieve an adequate “close-to-real-experience-prototype”.

1) *Visual Feedback:* Two types of visual feedback were implemented: the robot itself and a signal light. The robot’s animations naturally reflected all “processed” voice commands. While the light acted as an optional modality to

signalize that a command was recognized and an action sequence was started.

2) *Force Feedback*: For a more realistic simulation experience force feedback is essential. Haptic feedback modalities support the credibility of virtual reality with an active interaction channel [6]. As the visual feedback was easily implemented using a game engine for this simulation, the force feedback modality required some special adaptation. To support that feature, the plaster board was used as both input and as output device. The robot's actions were reflected to the user by specific force feedback according to each action performed. One motor controlled the simulated movements of the robot such as lifting the plate. Further, the Wii remote was used to demonstrate the robot's action of fixing the board with a drill.

V. SIMULATION CONTEXT

The simulation scenario was realized in the TV studio of the University of Applied Sciences, Salzburg, Austria. This location offered sufficient space and technical equipment to enable a credible setting. For the visual part two back projection screens were used. The primary screen showed the main interaction area that measured four meters in horizontal and three meters in vertical direction. The size of the screen reflected the common room height of a construction site. This interaction area was set up as an isolated environment to ensure an interaction without disturbances. Therefore, the primary screen bordered the real part of the scene in one direction. The second screen expanded the interaction area with a side view of the actual construction scene supporting the look and feel of a real room (see figure 3). This technique is similar to common virtual reality settings such as "the cave" [3].

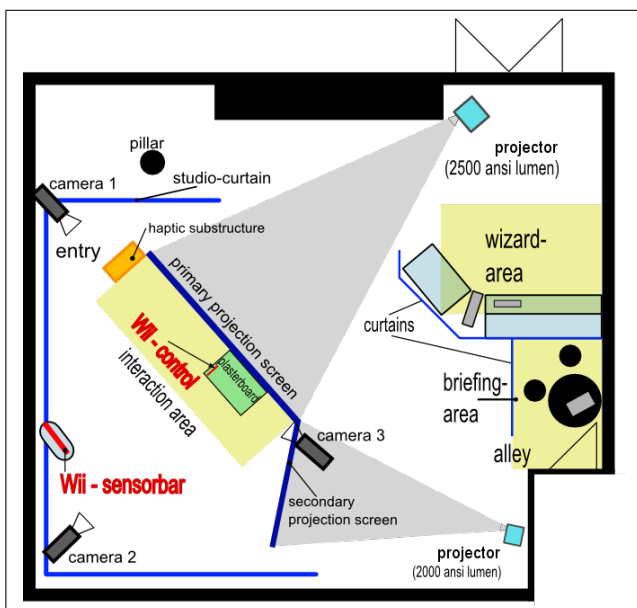


Fig. 3. Studio Set-up

To complete the construction site setting as an enclosed room, we used black curtains at the back of the interaction

area. These curtains did not affect the interaction experience as they were out of sight, behind the test participants. Thus, the interaction area was protected from external distractions and the test participants could focus solely on the task itself. Another advantage of the TV studio was the lighting equipment as working with projectors heavily depends on the lighting of the surrounding. To create a coherent environment we used the local equipment to dim the light according to both projectors' illumination intensity. To complete the whole setting, real construction site sounds were played in the background.

VI. PROOF-OF-CONCEPT USER STUDY

A. Study Setting

We conducted a user study to prove the feasibility of the proposed WOz set-up. The user study was based on a single task: The user should mount the plaster board together with the robot based on the action sequences presented in section III. The WOz set-up included all four necessary aspects: (1) The experimental set-up, consisting of four experimental conditions (Condition 0: Interaction without feedback; Condition 1: Interaction with visual feedback (blinking light showing that the robot understood the command); Condition 2: Interaction with haptic feedback; Condition 3: Interaction with visual and haptic feedback in combination). The natural speech interaction was simulated by the wizard. Therefore, the participants received five predefined verbal commands and were advised that the robot does not react on any other commands. The wizard listened to the participant and operated the actions of the robot like the following:

- 1) "Come to the board": The wizard started the action sequence "Walk to the board".
- 2) "Lift the board": The wizard started the action sequence "Grab board".
- 3) "Carry the board": The wizard started the action sequence "Carry board".
- 4) "Tilt the board": The wizard started the action sequence "Tilt plate".
- 5) "Screw plate": The wizard started the action sequence "Screw plate".

In the case that the person performing the wizard did not understand the verbal command or the participant did not give the exact word order, the participant was advised by the experimenter, who guided the participant through the study, to repeat the command. Experimenter and wizard were different persons. (2) To observe the user interaction and capture the data the scenario included, a set of microphones and two cameras, and a researcher additionally took notes during all tests. (3) To understand and measure if our WOz approach had sufficient interaction details and realism to evaluate user experience aspects, we distributed the AttrakDiff questionnaire [8] to the participants. The AttrakDiff is a questionnaire to measure the hedonic and pragmatic quality of an interactive system by numerous antithetic word-pairs, e.g. "disagreeable - likable". All items have to be graduated by the participants on a 7-point scale from the

negative word pole (-3) to the positive word pole (+3). In the analysis all items of this questionnaire are summarized into four scales: pragmatic qualities (PQ), hedonic qualities - identification (HQ-I), hedonic qualities - simulation (HQ-S), and attractiveness (ATT); a detailed description of these factors can be found in [8]. (4) The wizard was supported by a software allowing to trigger the five different actions the robot needs to perform in order to fulfill this task.

B. Results

24 participants took part in the study, counterbalanced in age, gender, and experimental condition. Eleven participants carried out the task successfully, but did not follow the ideal way in terms of minimum number of steps. Ten participants completed the task successfully, but with errors during single action sequences (e.g. wrong command, command given before the robot finished the previous action etc.). Only two participants needed a hint how to complete the task and only one participant aborted the task.

The results of the user experience values for the four experimental conditions showed that the users perceived the various forms of feedback differently. Significant differences could be revealed for the HQ-S scale ($F(3,20) = 3.20, p < .05$) and the ATT scale ($F(3,20) = 3.43, p < .01$). A post-hoc test (LSD) showed that condition 3 (interaction with visual and haptic feedback in combination) was perceived significantly better in the hedonic quality of simulation than all other conditions. Furthermore, the attractiveness was significantly better rated in condition 3 than in condition 0 (interaction without feedback) and condition 1 (interaction with visual feedback); it was also rated better than in condition 2 (interaction with haptic feedback), but this difference was not statistically significant. Similarly, a significant effect could be revealed for the overall scale of the AttrakDiff questionnaire, as condition 3 was rated better than the conditions 0, 1 and 2, but only for condition 0 and 1 the difference was statistically significant ($F(3,20) = 3.39, p < .05$). Based on the results of the AttrakDiff questionnaire it becomes clear that the different interaction techniques were presented realistic enough to allow the users to judge the user experience of the different interaction techniques. A mixed-reality WOz approach thus allows to prototype a system and to evaluate UX factors at early development (design) stages of a robot. The results of the user experience evaluation might not be generalizable for the final product or robot, but this type of study provides evidence for early design decisions in terms of user experience. For the study above the design recommendation for improving user experience when co-working with a humanoid robot would be to support the interaction technique with visual and haptic feedback. All participants also stated in the final interview that they perceived the WOz interaction technique prototype as sufficiently well designed to be able to judge the attractiveness and user experience.

C. Lessons Learned

The goal of this user study was to prove the feasibility of the proposed WOz set-up for evaluating user experience.

Considering our experiences we recognized the following issues as crucial in order to successfully evaluate user experience of multimodal interaction techniques in Human-Robot Interaction using a Wizard of Oz approach:

- 1) Evaluating user experience of human-robot interaction is possible, but a high fidelity mixed-reality prototype is necessary to allow a high degree of realism.
- 2) The pre-study showed that the various forms of interaction techniques were perceived differently in terms of user experience. The high fidelity prototype thus allowed to investigate different forms of interaction techniques in terms of user experience. The findings might not be generalizable for the final robot, but they allow to argue for one of the interaction techniques (if the goal is to improve user experience).
- 3) A mixed-reality approach including haptic feedback gives the user the feeling of “really” interacting with the robot. From a technical perspective the set-up for the haptic feedback needs careful preparation and additionally software (to allow to interpret the information coming from the Wii remote control).
- 4) From the technical perspective we found that participants wearing glasses had problems to focus on details in the projections. A projector with 1600 x 1200 pixel and a light intensity of 3000 ANSI lumen could probably solve this issue.

VII. CONCLUSION

To enable the evaluation of user experience we propose a high fidelity mixed-reality WOz set-up. Based on an experimental pre-study we have learned that a WOz set-up allows the evaluation of user experience of Human-Robot Interaction for collaborative tasks. Based on the experimental pre-study we can conclude that from a methodological perspective the WOz study can be helpful to investigate user experience, while it reduces the overall development costs for the (humanoid) robot. However, the WOz set-up is not trivial, as it needs knowledge in games programming, usage of augmented reality equipment, and additionally requires software to allow the wizard to control the tasks conducted during the experiment. As speech was perceived quite positive in terms of user experience we want to investigate possible influences of the (perfectly working) wizard compared to a speech recognition component. Future work will be the combination of a high fidelity prototype with a speech recognition component to investigate this possible influence on the perceived user experience of the interaction technique.

VIII. ACKNOWLEDGMENTS

The authors would like to thank all researchers supporting the prototype development, above all Michael Lankes and Thomas Mirlacher. This work is supported in part within the European Commission as part of the Robot@CWE project, see also www.robot-at-cwe.eu. The authors like to thank all partners from the project and gratefully acknowledge the collaboration with the researchers from CNRS-AIST JRL supporting us with the HRP-2 model.

REFERENCES

- [1] Christoph Bartneck and Jun Hu. Rapid prototyping for interactive robots. In *The 8th Conference on Intelligent Autonomous Systems (IAS-8)*, pages 136–145. IOS press, 2004.
- [2] Marion Buchenau and Jane Fulton Suri. Experience prototyping. In *DIS '00: Proceedings of the 3rd conference on Designing interactive systems*, pages 424–433, New York, NY, USA, 2000. ACM.
- [3] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142, New York, NY, USA, 1993. ACM.
- [4] Kerstin Dautenhahn. Methodology and themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, 4(1):103–108, 2007.
- [5] Ylva Fernaeus, Sara Ljungblad, Mattias Jacobsson, and Alex Taylor. Where third wave hci meets hri: report from a workshop on user-centred design of robots. In *HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 293–294, New York, NY, USA, 2009. ACM.
- [6] Mashhuda Glencross, Alan G. Chalmers, Ming C. Lin, Miguel A. Otaduy, and Diego Gutierrez. Exploiting perception in high-fidelity virtual environments. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*, page 1, New York, NY, USA, 2006. ACM.
- [7] Aders Green, Helge Hüttenrauch, and Kerstin Severinson Eklundh. Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration. In *Proc. IEEE Int. Workshop on Robot and Human Interactive Communication*, 2004.
- [8] Marc Hassenzahl. The thing and i: understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A. F. Monk, and P. C. Wright, editors, *Funology. From Usability to Enjoyment*, pages 31–42. Kluwer, Dordrecht, 2003.
- [9] Takayuki Kanda, Masayuki Kamasima, Michita Imai, Tetsuo Ono, Daisuke Sakamoto, Hiroshi Ishiguro, and Yuichiro Anzai. A humanoid robot that pretends to listen to route guidance from a human. *Auton. Robots*, 22(1):87–100, 2007.
- [10] Minkyung Lee and Mark Billinghurst. A wizard of oz study for an ar multimodal interface. In *ICMI*, pages 249–256, 2008.
- [11] Daniel Salber and Jolle Coutaz. Applying the wizard of oz technique to the study of multimodal systems. In *East-West International Conference on Human-Computer Interaction: Proceedings of the EWHCI93*, pages 55–67. Intl. Centre for Scientific And Technical Information, 1993.
- [12] Astrid Weiss, Regina Bernhaupt, Michael Lankes, and Manfred Tschelligi. The usus evaluation framework for human-robot interaction. In *AISB2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction*, pages 158–165, Edinburgh, Scotland, 8-9 April 2009. SSAISB (ISBN - 190295680X).
- [13] Regina Bernhaupt. *Evaluating User Experience in Games*. Springer, London, 2010.
- [14] ISO 9241-11. *Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability*. International Organization for Standardization, 1998.
- [15] Regan L. Mandryk, M. Stella Atkins, and Kori M. Inkpen. A continuous and objective evaluation of emotional experience with interactive play environments. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1027–1036, New York, NY, USA, 2006. ACM.