

Developing an usability test to evaluate the use of augmented reality to improve the first interaction with a product

Deise Albertazzi^{a,*}, Maria Lucia Okimoto^b and Marcelo Gitirana Gomes Ferreira^c

^{a, b} *Design Department, Federal University of Paraná, Edifício Dom Pedro I, Rua General Carneiro 460, 8 andar, Curitiba, Paraná, Brazil*

^c *Design Department, Santa Catarina State University, Avenida Madre Benvenutta 2007, Florianópolis, Santa Catarina, Brazil*

Abstract. Developing an usability test involves defining the metrics that are being evaluated, the methods used and the right tasks to be performed to achieve the objectives from the usability test. This article presents the development of an usability test aimed to evaluate the experience while first interacting with a product by using the augmented reality technology. This usability test is part of a bigger research that aims to evaluate if the use of augmented reality helps the process of learning how to use a new product. The paper starts with a literature review of related works, explaining many of the tools developed for evaluating usability tests. This section is followed by a methodological explain and the description of the developed usability test itself. The conclusion presents suggestions of further improvements.

Keywords: usability test, augmented reality, metrics, evaluation, product

1. Introduction

Usability tests are important resources to help evaluating product and systems. To obtain all the gains that an usability test may provide, it is important its development by a very accurate process. Planning an usability test includes knowing and defining the test objectives, the research questions, the participants characteristics, the methods, the task, the environment, equipments and logistics, the moderation rules, the gathering and evaluation of data and the results presentation [6].

Before a new technology is implemented to the general public, the usability tests help to understand its acceptability and provide useful feedbacks for fixing possible problems. Using augmented reality to improve the product usability is an example. Augmented reality (AR) is a technology aimed on providing the feeling that virtual elements are in the real world [3]. This concept is understandable when ana-

lyzing the continuum from Milgram, Takemura, Utsumi, & Kishino (1994), that shows the AR on a context that involves the real world at one side and the virtual world on the other side. The AR is placed in the middle, being a mixed reality that uses both real and virtual elements. In the moment you add one more element – the virtual element – on the reality, it is not predictable how the user will react.

There are some important questionnaires used to evaluate the product usability as the System Usability Scale (SUS) [2] and the Usefulness, Satisfaction and Ease of Use (USE) [4]. Both use Likert Scales to understand the feelings of the user after interacting with a product or system, by having the user answer the point scales that range from totally disagree to totally agree.

The SUS uses a Likert scale with ten phrases answered after the interaction between the user and a product or system and before any debate about it. The questionnaire results in an usability score be-

* Corresponding author. E-mail: dalbertazzig@gmail.com

tween zero and ten. The USE presents Likert scales from the following groups: usefulness, ease of use, ease of learning and satisfaction. It is possible to use questions from the specific groups to evaluate specific metrics. Usability tests are valuable to measure metrics such as performance, behavioral and psychological, combined and comparative, usability-issues based, and auto-reported metrics [7].

The PRISMA is a prototype of a binocular that uses augmented reality to improve the touristic experience [1]. During the research, it was gathered both quantitative and qualitative data from the usability test. The quantitative part aimed on evaluating usability and ease of use. The qualitative used the observation method to evaluate the user interaction, choices and movements.

The development of an usability test starts with the definition of the goals, researched questions, profile of the participants, methods, tasks list, environment, equipments and logistics, moderation rules, data collection and evaluation and results presentation [6].

Usability tests are fundamental tools to evaluate a new technology. This way, this paper explores the development of an usability test aimed on evaluating the use of augmented reality to improve the product usability.

2. Method

The first step to develop the test was to understand its context by researching previous works. The usability test from PRISMA revealed one of the possibilities to gather quantitative and qualitative data when evaluating the augmented reality technology. The next step was to understand the specific context of the test: goals, variables, metrics and participants.

The goal was to evaluate if the use of augmented reality helps the user as he interacts with a product for the first time. The independent variable was the way the user learns how to use the product. Four groups were evaluated: by direct manipulation, instructional manual, instructional video and augmented reality. The dependent variable was the usability, subdivided in the following metrics: task time, task success, errors, learnability, ease of use and satisfaction. The control variable was the technological familiarity with this kind of product: it was required that the participants have used a vacuum cleaner before. The factor test was the participant age: participants with less than 30 and 30 or more years old.

According to the heuristic evaluation [5], applying an usability test with 10 specialist participants results in discovering more than 90% of the usability issues that might be discovered. This way, the total number of participants is 40, 10 from each group and each group being subdivided in 2: participants with less and more than 30 years old, as illustrated in the Figure 1.

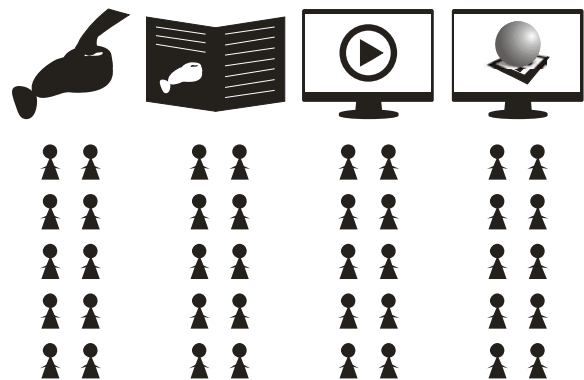


Figure 1
Schema of the participants

To be able to measure the required metrics, it was needed to develop a set of tasks to be performed by the participants of the usability test and a specific tool to evaluate these metrics. The developed task involves all the functions from the tested product - a portable vacuum cleaner. The three major tasks were: aspirating, emptying and assembling. Each task was divided in sub-tasks, as followed:

- Aspirating: turning on the vacuum cleaner; aspirating the paper balls (see Figure 2).
- Emptying: turning off the vacuum cleaner; opening the lid; disassembling the filter; take the paper balls away.
- Assembling: assembling the filter; assembling the lid.

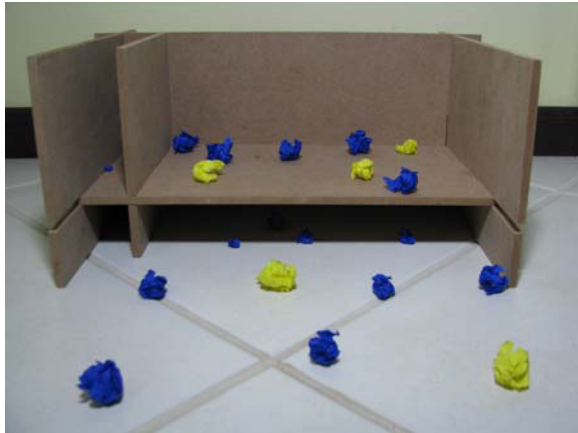


Figure 2:
The object designed for the test

Two instruments were developed to help the gathering of data. After performing the tasks, the users were asked to answer a 12 Likert scales questionnaire. This questionnaire was developed by using four sentences from each group of the USE questionnaire [4]. The questionnaire results on an usability score as described by the SUS questionnaire [2]. This questionnaire was used for gathering learnability, ease of use and satisfaction metrics data. The second instrument was used by the moderator to fill task-time, task-success and errors.

The pilot test task from this research was easy and fast to perform, not being able to provide the important feedback needed for the researches. Also, the questionnaire given to the participants failed while not mixing positive and negative sentences in the Likert scales. If the participant of the usability test had a great or a bad experience, his tendency is to answer positively or negatively most of the questions. If a questionnaire has only positive sentences, as 'I felt good using the product' and 'it is easy to use this product', the tendency is to answer all the sentences as if they were the same. Otherwise, if there are negative and positive sentences mixed as 'I felt good using the product' and 'it is difficult to use this product', the participant needs to pay attention to correctly answer all the sentences. The way de Likert scales were first developed made the participant choose only points in the left or in the right according to the experience they had.

3. Usability test

In the developed task the participant had to "clean" the blue paper balls from an object and not aspire the yellow ones (see Figure 3-a). The easiest way to clean the paper balls required the use of three different tools from the vacuum cleaner. After that, the participant had to open the vacuum cleaner and clean the paper balls from it (see Figure 3-b). In the end of the test the participant had to assemble the vacuum cleaner again (see Figure 3-c).



Figure 3
Participants performing the task

Just around 25 percent of the tests were performed already, but the early results indicate the tendency of making errors in three moments. The first one occurs when opening the lid to clean the vacuum cleaner by not removing it from the electrical outlet. The second error is the moment the participant tries to clean the filter not opening it, just removing the paper balls from the hole from it the paper balls entered. The third error occurs in the moment the participant assembles the filter, sometimes spending more time to perform this task than to perform all the others.

4. Conclusion

When developing an usability test for a new technology, an easy task prevent both the user and the moderator to find all the possible usability issues. It is important to explore all the functions in a partially difficult situation. The aim is not to force an error, but to let an error appear by simulating something not so easy and possible to occur in the regular use of the product. Also, a questionnaire with Likert scales should mix the questions and provide both positive and negative questions, in order to do not let the user automatically mark just the left or the right points because he liked or disliked the product. The participant should pay attention and read carefully each question to answer it. It is recommended that all usability tests are recorded using video cameras, as the actions occur really fast and it is not possible to obtain all the data and correct task-time just by watching the participant perform the task.

The initial results from the tests reinforce the need of evaluating products, even when a product is known as a simple product. Though most participants describe the portable vacuum cleaner as a product easy to use, none could perform the tasks without making an error yet. Without an usability test, an "easy to use product" or a "new interesting technology" might fail without the reason being discovered.

References

- [1] A. Alzua-Sorzabal, M. T. Linaza and M. Abad, An Experimental Usability Study for Augmented Reality technologies in the Tourist Sector, *Information and Communication Technologies in the Tourist Sector* 6 (2007), 231-242.
- [2] J. Brooke, A quick and dirty usability scale, in: *Usability evaluation in industry*, P. W. Jordan, B. Weerdmeester, A. Thomas and I. L. Mclelland, eds., Taylor and Francis, London, 1996.
- [3] S. Cawood and M. Fiala, *Augmented reality: a practical guide*, Raleigh, Pragmatic Bookshelf, 2008.
- [4] A. M. Lund, Measuring usability with the USE questionnaire, *Usability & User Experience* 8, (2001).
- [5] J. Nielsen, Finding usability problems through heuristic evaluation, *Proceedings of the SIGCHI conference on Human factors in computing systems*, (1992), 373-380.
- [6] J. Rubin and D. Chisnell, *Handbook of Usability Testing: How to plan, design and conduct effective tests*, Indianapolis, Wiley, 2008.
- [7] T. Tullis and B. Albert, *Measuring the user experience: Collecting, Analyzing, and Presenting Usability Metrics*, Burlington, Morgan Kaufmann, 2008.