Usability evaluation in Virtual Environments through empirical studies involving users

Beatriz Sousa Santos[1,2], Paulo Dias[1,2], Paulo Santos[1], Samuel Silva[1,2], Carlos Ferreira[3,4]

1- DETI- Universidade de Aveiro, Portugal
2- IEETA - Universidade de Aveiro, Portugal
3- DEGEI- Universidade de Aveiro, Portugal
4- CIO- Universidade de Lisboa, Portugal

ABSTRACT

This paper briefly justifies the relevance of empirical studies involving users (either controlled or exploratory experiments) as methods that should be used to study and evaluate usability in Virtual Environments. The paper also briefly presents some examples of such experiments that have been performed in order to study user performance while navigating in a Virtual Environment in different platforms.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: *Evaluation/methodology*

Keywords

Virtual Environments, usability, evaluation, user studies

1. INTRODUCTION

As new technologies emerge allowing systems that are easy to use at a smaller cost, Virtual Environments (VEs) are entering new domains such as tourism, business, and cultural heritage. Nevertheless, they still present many challenges to Human-Computer Interaction. The range of different interfaces which might be experienced, the many behaviors which might be exhibited, have made it difficult to understand participants' performance within VEs, and provide coherent guidance for designers [15]. Moreover, empirical studies on human factors are unusually difficult to plan and carry out, given the large number of variables to control and costs involved [16]. Nevertheless, usability is at least as necessary for VEs to reach their full potential, as to any other type of interactive system and there is a need for guidelines and background information about the added value or appropriateness of alternative solutions.

Having this in mind, we have been performing studies [4,12,13] to assess users' performance, satisfaction and comfort while carrying out a set of (mainly navigation) tasks in a Virtual Environment using several platforms. These platforms were a common desktop, a setup using a Head-Mounted-Display (HMD) with an orientation sensor, and a modified desktop where the image is projected on an ordinary screen.

Some of these studies were controlled experiments meant to answer through hypothesis testing simple questions that had been identified as relevant in previous work. Other studies were more informal and exploratory, performed in order to yield ideas for what should be our next questions.

Throughout controlled experiments user performance data were automatically logged while using the VE, and satisfaction and comfort data were obtained afterwards through a questionnaire and interviews. In the scope of exploratory studies, users were asked about their preferences concerning some environment parameters deemed relevant for their satisfaction or performance (as ambient illumination and head rotation angle gain) while using the various setups. We present these studies as an illustration of the way we have been using these methods in order to obtain some more insight into the issues that are relevant to the usability of our VE.

2. USABILITY EVALUATION IN VEs

Several alternatives to evaluate usability in Virtual Environment exist, each having its applicability, advantages, and disadvantages [2]. Some methods widely used to evaluate more traditional interfaces, as Graphical User Interfaces (GUIs) [5, 10], seem readily adaptable and have been used to evaluate user interfaces in VEs (e.g. observation); others, can be more difficult to use, as heuristic evaluation, which involves a list of heuristics fine-tuned to the situation. General evaluation methods such as, controlled experiments, observation techniques, and questionnaires have been used in usability evaluation in VEs; yet, the fact that there are some important differences between evaluation of VE user interfaces and traditional GUIs should not be forgotten [1,2]. Some of these differences concern, for instance, physical environmental issues (e.g. while using a HMD, users may be standing, not sitting, and cannot see the surrounding physical

world), and others relate to the fact that more than any other current computer paradigm, VEs involve user's senses and body in the task.

Despite the growing interest in usability related research in the VE community, not as many papers concerning usability evaluation exist, as compared to papers proposing new methods, techniques or systems. For instance, in a research recently conducted, we were able to find only a few studies comparing user performance while using VEs in desktops and systems including a HMD [12,13]. Analyzing these studies, it can be observed that controlled experiments involving users have been the most used evaluation method, complemented in some cases with a questionnaire.

User studies have been considered an important method in other contexts, as Scientific Visualizations and Augmented Reality [6,9]. We believe that they can, likewise, contribute to optimize VEs informing their design within a usability engineering approach; however they can also be used to compare alternatives, validate solutions, and more fundamentally help seeking insight into why a particular solution is effective, thus allowing to establish design guidelines.

3. EXPERIMENTS INVOLVING USERS

As mentioned before, we have been conducting a series of studies involving users in different VE platforms (a desktop, a setup using a HMD and a setup where the image was projected on a screen) in order to compare their usability. All experiments were based in the same virtual maze where users had to navigate for a fixed amount of time in order to catch objects [12,13]. The results obtained from automatically logged data, as well as observation of (and interviews with) participants raised a number of questions concerning the usability of the different setups that motivated a new set of studies. As main questions in this set of studies we defined the following:

1. Is user performance improved even by a small amount of training with the HMD?
2. Is user performance influenced by the input device (e.g. mouse vs. joystick) when using the HMD?
3. What are the users' preferences concerning the head rotation gain and posture (standing vs. seating) when using the HMD?
4. What are the users' preferences concerning ambient illumination and distance while viewing the image projected on a screen?

These questions are different in nature; hence they had to be tackled using different methods. The first two correspond to explicit precise hypothesis and thus can be dealt with the help of confirmatory studies. Questions 3 and 4 are less clear, vaguer, and might seem informal and aimless nevertheless they correspond to a first phase of a research project, having the potential to yield new explicit questions that can later be tested in controlled experiments. Exploratory studies are suitable to this first phase [3].

Taking into consideration these differences, we used controlled experiments (involving hypothesis, input and output variables, as well as statistical analysis) to search for an answer to questions 1 and 2, and exploratory studies in order to study the others. In the following sections we briefly describe each study.

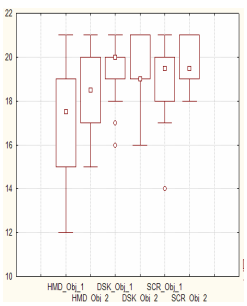
3.1 Influence of training with the HMD

In our previous studies, we noticed that users had generally performed better while using the desktop; however, the difference to the HMD setup was not very large, and thus we hypothesized this difference might be a consequence of the fact that the majority of our users had much experience using a desktop and none using a HMD. To get some more insight into this matter we decided to assess the performance of a group of users before and after training with the HMD. This study was performed as a controlled experiment. Our null hypothesis was that users would not improve their performance after training for a small amount of time (4 periods of 5 minutes each); an in-group experimental design was used; setup (HMD vs. desktop, vs. screen) was the input variable and performance the output variable [10]. Performance was measured through the number of caught objects, velocity attained and number of collisions with the walls.

This experiment was run along several days, with a group of 15 volunteer students that were trained using the HMD. In the first week their "base performance" was measured in the same conditions as in a previous study [12,13], and after the training was complete, performance was measured again maintaining the experimental conditions. Meanwhile, they trained in the HMD setup using a different maze as we didn't want them to form a mental map which could afterwards influence the results.

Performance data collected in both moments (before and after training) were analyzed using Exploratory Data Analysis [8], and some tests were performed using STATISTICA [14]. Figures 1 and 2 show box-plots corresponding to objects caught, number of collisions and velocity attained by users while using all setups (HMD, DSK, SCR), before and after training (1 and 2). We notice a slight increase in the median number of objects caught by the users after they had been training with the HMD (HMD_OBJ_2), as compared to the number caught before any training (HMD_OBJ_1), as well as a slight decrease in the number of collisions (see HMD_COL1 and HMD_COL2). However, using Wilcoxon Matched Pairs Tests [7], the difference concerning the number of caught objects was found not significant (p-value=0,14), while the difference concerning the number of collisions was found significant (p-value=0,015). Analyzing figure 2, we also notice a slight increase in velocity in all setups; though the Wilcoxon Matched Pairs Tests found significant only the difference concerning the HMD (p-value=0,020).

The results of this study must be considered with some caution, since it has some limitations: the small amount of training given to users, the small number of users, the small number of objects to catch and even the specific characteristics of the maze used in the VE, might have influenced the results. Thus, this experiment should continue, either training more users, more time or with other mazes.



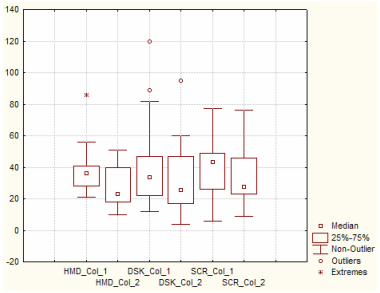


Figure 1- Box-plots corresponding to the number of objects caught and collisions while using the three setups, before (1) and after (2) training

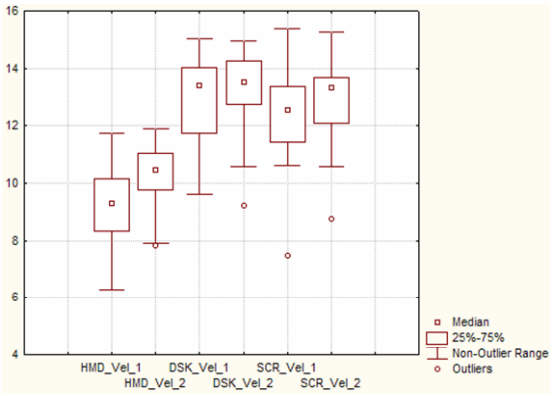


Figure 2- Box-plots corresponding to velocity attained by users while using the three setups, before (1) and after (2) training

3.2 Mouse vs. joystick while using the HMD

In our previous studies the input device in the setup using the HMD was a 2 button device - a mouse - for moving forward and backward; however, many users commented that a joystick might be a better alternative. This observation motivated another controlled experiment where we started from the following null hypothesis: users would have similar performance whether using a mouse or a joystick. A between-group experimental design was used; input device (mouse vs. joystick) was the input variable and performance the output variable, again measured through number of caught objects, velocity and number of collisions with the walls. This time we had the collaboration of a group of 21 students, which were divided in two groups: 10 used the mouse and 11 used the joystick. In order to have a "base performance", all participants used also the desktop. The same type of data analysis was performed. Figure 3 shows the box-plots corresponding to the number of objects caught by each group as using the desktop (G1_d, G2_d), and the HMD with the mouse (G1_m) or the joystick (G2_j). We observe that users caught always a higher number of objects with the desktop. Since we had two groups of users we tested if the difference between their performance with the desktop was significant, and found it was not using a Mann Whitney U Test (p-value=0,31). Then, testing the difference between number of objects caught by the two groups with the HMD (either using the mouse or the joystick), we found again that it was not significant (p-value=0,17). This result came as a surprise since we were convinced users would have a better performance using the joystick (the comments and preferences expressed by many of them had given us that impression). Again, this might be due to the relatively small number of users (10+11).

Nevertheless, the results obtained confirmed again the fact that users generally perform better while using the desktop, being consistent with all previous experiments. In fact, the difference in number of objects caught by each group using the HMD (either with the mouse or the joystick), and the desktop was established as significant through Wilcoxon Matched Pairs Tests (p-value= 0,00009, p-value=0,00004, respectively).

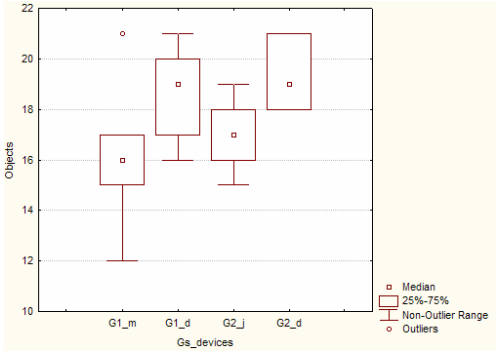


Figure 3- Box-plots corresponding to the number of objects caught and collisions while using the mouse and the joystick

3.3 Preferences with the HMD

Based on observation of users while they were using the HMD, as well as, on their opinions, comments and suggestions, we had become aware of the fact that several parameters and configurations of the setup might be relevant to user satisfaction and comfort, if not to performance. In order to obtain some more insight into this issue, and since running controlled experiments is quite time consuming and onerous in term of users and other resources, we decided to run first a simple exploratory study concerning a few parameters, namely the system's head rotation gain and user posture (standing vs. sitting). This was done as a series of interviews where we asked 21 students to use the system for a while with several values of head rotation gain and in both postures, and let us know their preferences.

Posture	*N. users*
Seat	7
Standing	14
Total	21

Table 1 – Users preferences concerning posture with the HMD setup

As can be observed in table 1, two thirds of the users preferred to stand saying that the experience would feel more natural. The other users preferred to seat; however a few mentioned that head rotation gain should be adjusted to this posture, higher than 1 (e.g. a 45º head rotation should correspond to a 90º rotation in the VE would mean gain=2). To study preferences concerning head rotation gain, we asked the same users to use the system with five different values of gain (1, 3, 4.5, 6, 9). Table 2 shows the results. Most users preferred intermediate values (between 3 and 6). It is interesting to notice that none preferred the value 1, (i.e. that the rotation of their heads would result in the same rotation angle in the VE as in the physical world). This might be related to the fact that cables were often entangled around them as they turned, while exploring the maze in this condition, which is in agreement with information collected in previous studies.

HRGain	1	3	4,5	6	9
N. users	0	11	3	6	1

Table 2 – Users preferences concerning head rotation gain with the HMD

3.4 Preferences with the screen projection

We had also become aware that there might be different preferences concerning ambient illumination and distance to the

screen in this setup., and thus we asked 17 students about their preferences after letting them use the setup for a while.

Most users (15) preferred to use the VE without ambient light, which might be due to the fact that this gives an enhanced feeling of immersion (everything else disappears, and the users can better concentrate on their task). In order to collect preferences concerning distance to the screen, we had a chair placed 2m away from the screen and asked users to place it at their preferred distance. We observed that all users moved the chair less than 30 cm either nearer or further the screen.

4. DISCUSSION

The first two studies illustrate how controlled experiments with users can be applied to study usability issues in VEs. These are simple examples that produced not entirely conclusive results, probably because we were not able to gather a representative enough sample; still, they enhanced our understanding of how people use our VE in specific circumstances and provided us with some more ideas on how to further our research. Moreover, they confirmed previous results concerning the user performance difference while using HMD and desktop.

Controlled experiments involving users are in our opinion adequate if we aim at searching for insight, establishing guidelines for the development of more usable VEs, or performing summative evaluation. Nevertheless a word of caution is needed: designing controlled experiments is a complex procedure and a statistician should be consulted to avoid unpleasant surprises at the end, and guarantee that the collected data can be treated statistically in order to test the chosen hypothesis.

If we just want to explore unknown situations, we should use exploratory studies, which have the potential to yield new explicit questions that can later be tested in controlled experiments, which will motivate new questions, and so on.

5. CONCLUSIONS

The problem of studying and evaluating Virtual Environments concerning their usability is a challenge to the Human-Computer Interaction community. It seems that not all well established methods for traditional UI evaluation can be easily adapted; yet we believe methods involving users are crucial since they don't use a model, but "the real thing"; moreover, they will be applicable to Reality-Based Interfaces in general. On the other hand, they may be time very consuming and otherwise onerous, in their more controlled forms; however, we also believe that in a variety of situations, exploratory studies are adequate.

This problem urgently needs attention by the VE community, and we believe all the efforts toward tackling it are valid, even if they are crude at first.

6. REFERENCES

1. Bowman, D, Kruijff, E, LaViola, Jr J, Poupyrev, I. 2001. An introduction to 3D user interfaces design. Presence: Teleoperators and VR, vol.10, 1, pp. 96-108
2. Bowman, D, Kruijff, E, LaViola, Jr J, Poupyrev, I. 2005. 3D User Interfaces: Theory and practice, Addison Wesley
3. Cohen, P., Empirical Methods for Artificial Intelligence, The MIT Press, 1995
4. Dias, P., Pimentel, A., Ferreira, C., Sousa Santos, B. 2007. Usability in Virtual and Augmented Environments: A Qualitative and Quantitative Study, Electronic Imaging 2007: Stereoscopic Displays and Virtual Reality Systems XIV, SPIE Vol. 6490, San José, USA, 64901I/1-I/11
5. Dix, A., Finlay, J., Abowd, G., Beale, R. 2004. Human-Computer Interaction. 3rd ed., Prentice Hall
6. Gabbard, J., Swan II, J. Usability Engineering for Augmented Reality: Employing user-based studies to inform design. 2008. IEEE Transactions on Visualization and Computer Graphics, vol. 18, 3, pp. 513-525
7. Hettmansperger, T., McKean, J. 1998. Robust Nonparametric Statistical Methods. Kendall´s Library of Statistics, Vol.5, Arnold
8. Hoaglin D, Mosteller, F, and Tukey J. 1983. Understanding Robust and Exploratory Data Analysis, John Wiley & Sons
9. Kosara, R., Healey, C., Interrante, V., Lidlaw, D., Ware, C. 2003. User Studies: Why, How, and When?, IEEE Computer Graphics and Applications, July/August, pp. 20-25
10. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T. 1994. Human-Computer Interaction, Addison Wesley
11. Siegl, H., Hanheide, M., Wrede, S., Pinz. 2007. A., An augmented reality human–computer interface for object localization in a cognitive vision system, Image and Vision Computing, vol. 25, pp. 1895–1903
12. Sousa-Santos, B., Dias, P., Pimentel, A., Baggerman, J.W., Ferreira, C., Silva, S., Madeira, J. 2008b. Head Mounted Display versus desktop for 3D Navigation in Virtual Reality: A User Study, Multimedia Tools and Applications (in press)
13. Sousa-Santos, B., Dias, P., Silva, S., Capucho, L., Salgado, N., Lino, F., Carvalho, V., Ferreira, C. 2008a. Usability Evaluation in Virtual Reality: A User Study Comparing Three Different Setups. Proceedings of Eurographics Symposium on Virtual Environments EGVE08 (Posters), B. Mohler, R. Van Liere(ed.), Holland (2008), 21-24
14. Statistica 6.0. http://www.statsoft.com (on-line Oct/2008)
15. Wilson, J. 2006a. Interaction with Virtual environments. Int. J. Human-Computer Studies, 64, pp.157
16. Wilson, J, D´Cruz, M. 2006b. Virtual and interactive environments for work of the future. Int. J. Human-Computer Studies, vol. 64, pp158-169