EGVE Symposium (2008)
B. Mohler and R. van Liere (Editors)

Usability Evaluation in Virtual Reality: A User Study Comparing Three Different Setups

B. Sousa Santos[1,2] P. Dias[1,2] S. Silva[1,2] L. Capucho[1] N. Salgado[1] F. Lino[1] V. Carvalho[1] and C. Ferreira[3,4]

[1] DETI – University of Aveiro, Aveiro, Portugal
[2] IEETA – University of Aveiro, Aveiro, Portugal
[3] DEGEI – University of Aveiro, Aveiro, Portugal
[4] CIO – University of Lisbon, Lisbon, Portugal

Abstract

We describe a user study comparing a low cost VR system using a Head-Mounted-Display (HMD) to a desktop and another setup where the image is projected on a screen. Eighteen participants played the same game in the three platforms. Results show that users generally did not like the setup using a screen and the best performances were obtained with the desktop configuration. This result could be due to the fact that most users were gamers used to the interaction through keyboard/mouse. Still, we noticed that user performance in the HMD setup was not dramatically worse and that users do not collide as often with walls.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces: Input Devices and Strategies

1. Introduction

With the development of new lightweight less expensive systems, Virtual Reality (VR) is entering many new areas. On the other hand VR does not necessarily mean immersive stereoscopic visualization as many emerging VR applications tend to be desktop based and not stereoscopic. Despite the growing interest, only a few usability tests and evaluations exist (see [RP04, QTHM06, PFS*08] for some recent examples) which might provide important guidelines for using and implementing VR systems.

While using a Virtual Environment (VE), before performing other tasks, users must often be able to navigate: they must determine their present location, find where they want to go, plan their route and follow it. Navigation is therefore one of the core tasks people perform in VEs [BKLP05]. Therefore, there is much interest within the field of VEs on how different forms of interaction and a variety of environmental characteristics might affect navigational learning.

The importance of navigation in VEs, as well as the interest in the differences in performance obtained in immersive and non-immersive VEs, lead us to compare the usability of a low cost VR system under development at our University (using a Head Mounted Device - HMD and a 3 Degree of Freedom (DOF) Head Tracker), to a desktop and a setup where the image is projected on a screen.

This comparison was performed through a user study to assess the performance and satisfaction of 18 users in all setups as they navigated through the VE. In this paper we shortly describe the experiment, as well as its main results.

2. Setups and Virtual Environment

Our VR system consists of stereo HMD i-glasses SVGA Pro with a resolution of 800×600 pixels, stereoscopic capabilities, 26^o Field of View, a frame rate of 60Hz or 120 Hz (corresponding to mono or stereo respectively), an orientation sensor (tracker) InterTrax 2 from InterSense with three degrees of freedom (DOF) and a PC with a nVidia Quadro FXGo 1400 graphics card. A two button device was used to allow forward and backward movement according to the user's head direction.

As mentioned, our study involved comparing user performance using this system and two other setups, a desktop and a setup where the image was projected on a screen using an ordinary projector and the interaction was performed through a keyboard and mouse as in the desktop.

Figure 1: *Setup using a HMD and a view of the used Virtual Environment.*

The desktop used in this study had a 19" Wide Screen monitor with a resolution of 800×600 pixels, and the screen was $1,4m \times 1,5m$ and placed in front of the user at approximately $1m$.

As VE we chose a maze since it was simple to build using existing tools and it provides a good way to test some navigation alternatives (other authors also used mazes in navigation experiments [RP04]). A gaming scenario can be made to incorporate several navigation tasks that users have to perform spontaneously (such as forward motion, cornering and navigating through doorways) while instilling a sense of competition between users. The game used in this experiment was adapted from an already existing game, that had been developed for a previous study [DPFSS07] and is based on a maze having very similar corridors in which users had to navigate in order to find 21 objects (see figure 1). These objects were floating at eye level and the users had only to collide with them receiving an audio feedback when they were successful.

3. Experiment

We asked for the collaboration of eighteen volunteer users (six women and twelve men) aged from 18 to 26, the great majority of whom were students from different departments of our University. None had experience in using VR systems and most had gaming experience.

After an initial presentation concerning the experiment, users started playing the game without any previous training. During the game, an observer was monitoring their performance and taking down relevant information.

We started from a simple hypothesis that the performance and satisfaction of users would be similar in all conditions: using the HMD, the screen and the desktop (that we will call H, S and D), which are the levels of our independent variable. As dependent variables we used user performance related variables. In our study we deemed experience in viewing 3D computer scenes, as well as in gaming as potentially influencing user performance and thus addressed these issues in the questionnaire we gave to users. A further possible secondary variable is previous experience with the game in another condition (D, S, H). This experience (and consequent learning) could increase user performance on the other condition and so we controlled the sequence in which users would experience the three conditions.

During the game, users had to navigate in the virtual maze to locate and collect as many objects as possible (out of 21 possible) for a period of 5 minutes. User performance was assessed via a set of automatically logged quantitative measures: number of collected objects, number of collisions with the walls, traveled distance, total gaming time and position in the maze over time. Also, some relevant information concerning users' behavior and performance was registered during the game by an observer (e.g., the number of times each user failed to catch an object in the first attempt, the number of complete turns and other difficulties).

After playing in all conditions (D, S and H), users were given a questionnaire with a few questions about their profile (as age, gender, profession, game playing habits, experience in using 3D), as well as about their satisfaction, preferences and opinions regarding the three playing modes.

We chose a within-subjects experimental design i.e., all subjects performed under the three conditions, due to the advantages of requiring less subjects, as well as reducing the effect of individual differences [DFAR04].

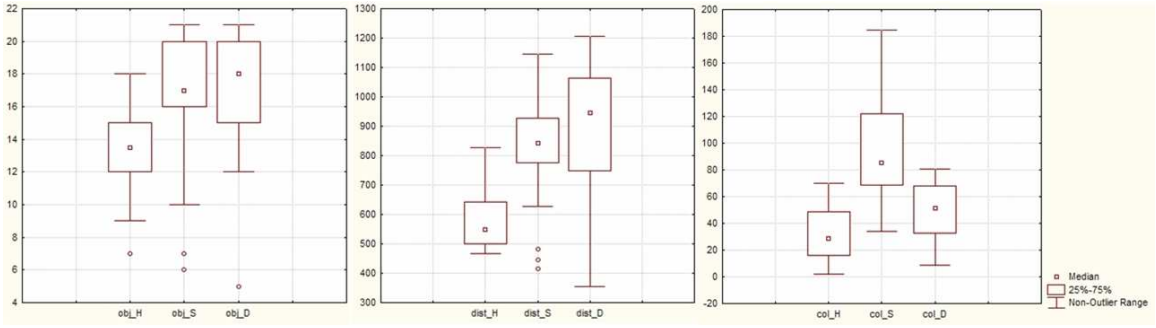
To avoid a possible bias on the results due to learning (as mentioned) or boredom, the order in which conditions were tackled was varied among users, as to have an equal number of all (six) possible sequences (e.g., D-S-H, D-H-S, etc.).

4. Results

In this section we present the main results obtained from the collected data. Due to the fairly small number of users, we mainly applied Exploratory Data Analysis techniques and non-parametric tests using Statistica [Sta08].

Figure 2 shows the box-plots [HMT83] corresponding to the data logged while the users were playing: collected objects, distance traveled and collisions with the wall. From these box-plots we get the general idea that user performance in conditions D and S is more similar while performance in condition H differs slightly.

Observing figure 2a) we see that the median number of objects caught by the users while playing in condition H $(13,5)$ was inferior to the objects caught in the other conditions (17 and 18 in S and D, respectively). Using Wilcoxon tests [Con99] to compare all possible combinations of condition pairs, (H,S), (S,D) and (H,D), we obtained the values $p = 0,007$, $p = 0,046$, $p = 0,004$, respectively (all



a) b) c)

Figure 2: *Box-plots of a) number of objects collected; b) distance traveled; and c) number of collisions for each condition.*

$p < 0,05$), meaning that the number of objects caught in the three conditions is significantly different.

In figure 2b) we notice that users traveled smaller distances in condition H, than in condition S, and condition D. Again using Wilcoxon tests to compare all possible combinations of condition pairs ((H,S), (H,D) and (S,D)), we obtained the values $p = 0,0005$, $p = 0.0004$ ($p < 0,05$) and $p = 0,08$ ($p > 0,05$), respectively, meaning that the differences of traveled distances in condition H and the other conditions are significant; however the distances traveled in S and D are not significantly different.

Observing figure 2c) we notice that the number of collisions is smaller in condition H , and higher in condition S. The median values for all conditions are significantly different, as found by Wilcoxon tests to all combinations of condition pairs (H,S), (H,D) and (S,D) ($p = 0,00002$, $p = 0.0008$, $p = 0,0002$, respectively, all $p < 0,05$).

To extend this analysis, and concerning user performance differences among the three conditions, we used Correspondence Analysis [HMT83], which shows that the three conditions differ mainly concerning the number of collisions and less concerning the number of objects caught, as can be seen in the factorial plane presented in figure 3, where the values of collected objects in all conditions are very near, and the values of collisions are far apart. This was confirmed through a Principal Component Analysis.

These results seem to imply that users had generally better performances using conditions D and S than condition H, as they collected more objects and traveled greater distances. However, they adopted a more careful strategy concerning collisions in condition H, which could be due to the fact that collisions might be more realistic while using the HMD.

In order to investigate the influence on users' performance of previous experience with the game in another condition (D, S, H), as well as boredom, identified as two possible secondary variables in this experiment, we studied the performance of the users categorized by the followed sequence

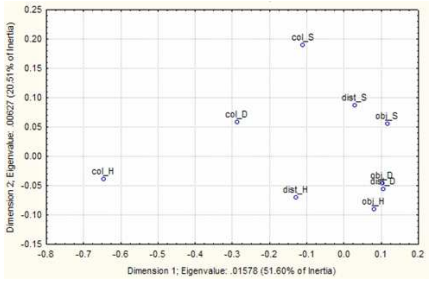


Figure 3: *Correspondence Analysis showing user performance in the three conditions.*

of conditions and verified no significant influence. Nevertheless, a word of caution is due, since this result was obtained with only three users per sequence

As previously mentioned, some relevant information concerning users' behavior and performance was registered by an observer during the game. Analyzing the data collected by all observers we notice that eleven users failed more when trying to catch objects while using the HMD. It is also apparent that most users failed less when using the desktop. This might be due to the fact that our users did not have any previous experience with our VR system but the great majority had gaming experience in desktops.

We observe much less users' distractions while using the screen setup when compared with the HMD and the desktop. This might mean that its large size and short distance had created a stronger immersion.

According to the answers given in the questionnaire, users preferred the desktop, closely followed by the HMD. The screen setup was only preferred by one user. After some analysis we concluded that this poor result observed for the screen setup had mainly to do with the fact that the screen was too near the user, at least for the type of application used where most users move around very fast.

	Desktop	Screen	HMD
Satisfaction	4	3	4
Difficulty	1	2,5	3

Table 1: *Median values of users' satisfaction and difficulty concerning all conditions (in a scale from 1 – very unsatisfied / very easy to 5 – very satisfied / very difficult).*

The median values of satisfaction and difficulty concerning the three conditions, as rated by users in a five level Likert scale (1 – very unsatisfied, 5 – very satisfied and 1 – very easy, 5 – very difficult), are shown in table 1. Clearly users generally found the desktop condition very easy, as well as satisfying; they also found the HMD satisfying, however not so easy to use; the screen condition was rated in between, either in satisfaction or in difficulty. It is interesting to notice that, even if only one user had preferred the screen condition, the overall user satisfaction was still positive.

In addition, we collected comfort data concerned with negative side effects on the user, specifically dizziness, seasickness and nausea. The overall results showed that the desktop was the condition where users were less affected by these negative side effects, and the screen was the condition where they felt worse. We believe this was the main reason why the screen was the less preferred condition and had the lower satisfaction results.

5. Discussion and Future Work

This paper describes a user study performed in order to compare the usability of three different VR setups, one based on a HMD, another based on a desktop and a third where the image is projected on a screen. This study was devised as a controlled experiment including observation, and a questionnaire given to the users in order to assess their opinion and satisfaction on the experience they had. The main results were:

- Users generally performed better using the desktop; however, the difference to the other setups is not very large (specially to the screen setup), and this difference could be a consequence of the fact that the majority of our users had gaming experience using a desktop and did not have any previous VR experience;
- Users were as satisfied with the HMD as with the desktop even though they had more difficulties with the HMD;
- The screen was the less preferred setup; however, user performance was similar to the desktop;
- Several users felt negative effects (dizziness, seasickness and nausea) when using the screen setup, which might be the main reason why it was the least preferred one.

Despite the relatively poor results of preference and negative effects obtained for the screen setup as compared to the others, we believe it is promising, at least for some applications, since it seems to create an interesting level of immersion using only a PC and a common projector (i.e., at a low cost); moreover, we believe that by adjusting the intensity level and distance to the user its usability may be much improved. Future experiments shall be performed to address this along with other issues such as the influence of previous experience using the VR equipment in user performance.

Acknowledgments

A word of gratitude is due to Jan-Willem Baggerman and Angela Pimentel for their collaboration in a previous experiment upon which this study was based and to the 18 participants for their collaboration.

References

[BKLP05] BOWMAN D., KRUIJFF E., LAVIOLA J. J., POUPYREV I.: *3D User Interfaces: Theory and practice*. Addison Wesley, 2005.

[Con99] CONOVER W. J.: *Practical non-parametric Statistics*, 3rd ed. John Willey & Sons, 1999.

[DFAR04] DIX A., FINLEY J., ABOWD G., RUSSEL B.: *Human Computer Interaction*, 3rd ed. Prentice Hall, 2004.

[DPFSS07] DIAS P., PIMENTEL A., FERREIRA C., SOUSA SANTOS B.: Usability in virtual and augmented environments: A qualitative and quantitative study. In *Electronic Imaging Conf., SPIE Vol. 6490, Stereoscopic Displays and VR Systems XIV* (2007), p. 64901I.

[HMT83] HOAGLIN D., MOSTELLER F., TUKEY J.: *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons, 1983.

[PFS*08] PRABHAT M., FORSBERG A., SLATER M., WHARTON K., KATZOURIN M.: A comparative study of desktop, fish tank and CAVE systems for the exploration of volume rendered confocal data sets. *IEEE Transactions, on Visualization and Computer Graphics (in press)* (2008).

[QTHM06] QI W., TAYLOR R., HEALEY C., MARTENS J.: A comparison of immersive HMD, fish tank VR and fish tank with haptics displays for volume visualization. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (2006), pp. 51–58.

[RP04] RUDDLE R., PÉRUCH P.: Effects of proprioceptive feedback and environmental characteristics on special learning in virtual environments. *Int. Journal of Human-Computer Studies 60* (2004), 299–326.

[Sta08] Statistica 6.0. *http://www.statsoft.com* ((online Jan/2008)).