Usability in Virtual and Augmented Environments: A qualitative and quantitative study

Paulo Dias¹,a,b, Angela Pimentel c, Carlos Ferreira c,d, Frank van Huussen e, Jan-Willem Baggerman e, Pauline van der Horst e, Joaquim Madeira a,b, Rafael Bidarra e, Beatriz Sousa Santos a,b

¹Departamento de Electrónica, Telecomunicações e Informática, Univ. de Aveiro, Aveiro, Portugal
²IEETA – Instituto de Engenharia Electrónica e Telemática de Aveiro, Aveiro, Portugal
³Departamento de Economia, Gestão e Engenharia Industrial, Univ. de Aveiro, Aveiro, Portugal
⁴CIO – Centro de Investigação Operacional, Universidade de Lisboa, Lisboa, Portugal
⁵Faculty EEMCS, Delft University of Technology, Delft, The Netherlands

ABSTRACT

Virtual and Augmented Reality are developing rapidly: there is a multitude of environments and experiments in several laboratories using from simple HMD (Head-Mounted Display) visualization to more complex and expensive 6-wall projection CAVEs, and other systems. Still, there is not yet a clear emerging technology in this area, nor commercial applications based on such a technology are used in large scale. In addition to the fact that this is a relatively recent technology, there is little work to validate the utility and usability of Virtual and Augmented Reality environments when compared with the traditional desktop set-up. However, usability evaluation is crucial in order to design better systems that respond to the users’ needs, as well as for identifying applications that might really gain from the use of such technologies.

This paper presents a preliminary usability evaluation of a low-cost Virtual and Augmented Reality environment under development at the University of Aveiro, Portugal. The objective is to assess the difference between a traditional desktop set-up and a Virtual/Augmented Reality system based on a stereo HMD. Two different studies were performed: the first one was qualitative and some feedback was obtained from domain experts who used an Augmented Reality set-up as well as a desktop in different data visualization scenarios. The second study consisted in a controlled experiment meant to compare users’ performances in a gaming scenario in a Virtual Reality environment and a desktop. The overall conclusion is that these technologies still have to overcome some hardware problems. However, for short periods of time and specific applications, Virtual and Augmented Reality seems to be a valid alternative since HMD interaction is intuitive and natural.

Keywords: Virtual Reality, Augment Reality, Usability, Evaluation

1. INTRODUCTION

Virtual Reality (VR) as well as Augmented Reality (AR) are important emerging technologies with applications in many different areas, such as automotive industry, architecture, aeronautics, medical science, education and entertainment.

In VR and AR, as for any other product or system, designers should consider usability issues: the ability to carry out tasks effectively, efficiently and with satisfaction by their intended users [1]. That is, design systems as simple and usable as possible. Although VR and AR have improved in the last few years, it is still necessary to define methods and parameters to test and evaluate the usability of such systems during their development and implementation cycle [2]. Although usability evaluation tools, methods, and techniques are well defined for 2D environments, there are relatively few for the evaluation of 3D Virtual and Augmented Reality systems [1].

¹ paulo.dias@det.ua.pt; phone +351 234 370 531; fax +351 234 370 545
It is a common idea that Virtual and Augmented Reality systems can facilitate the execution of tasks. But, if they are not well designed, users will face usability problems, such as motion sickness, difficulty of navigation or spatial disorientation [3][4]. Furthermore, there are other problems specific to these technologies: high cost, performance and availability of software, database management, input and output devices, data visualization, mathematical description and representation of physical processes. Due to all these limitations, many experts still find traditional desktop systems easier to use than virtual systems. Usability evaluation also faces another problem: few companies are interested in investing money and resources to evaluate usability; this may result in systems that are not adapted to intensive work and are used only due to innovation or curiosity without providing any real benefit.

The work described here is a first step towards evaluating the usability of a low-cost Virtual and Augmented Reality environment under development at the University of Aveiro [5]. This is a generic prototype for exploration of three-dimensional models and data that uses a stereo head-mounted display (HMD), a wireless video camera and an orientation sensor.

This environment can be used in two different modes: Augmented or Virtual Reality. In order to study the usability of both modes we have devised two different experiments corresponding to application areas we consider promising: scientific data visualization for Augmented Reality and games for Virtual Reality. Moreover, games can involve tasks (e.g., orientation and navigation in a virtual world or locating and catching objects) that can be used in other contexts where Virtual Reality can be useful such as training workers to perform real tasks.

According to the tasks, contexts and performance criteria used we can categorize user studies in abstract and application-specific [6]. Whereas the Augmented Reality user study was application-specific, since it involved users analysing their own data, we consider the Virtual Reality experiment as an abstract user study. In both experiments the participants performed as well equivalent tasks in a traditional desktop PC in order to compare their performance, satisfaction or opinion in both conditions.

In the following sections we describe the experiments and applications developed to test both modes. Results are then presented for the studies in Augmented Reality and Virtual Reality. Finally, some conclusions are drawn and possible directions of future work are presented.

2. TEST ENVIRONMENTS

The specific hardware used in our low-cost VR and AR environment consists of a stereo HMD i-glasses SVGA Pro with resolution of 800x600 pixels, an orientation sensor (tracker) InterTrax 2 from InterSense with three degrees of freedom (yaw, pitch and roll), and a wireless analog micro-camera with a CMOS sensor of approximately 330 lines, working at 1.2GHz. Given the small dimensions and weight of the camera, it was possible to mount it directly at eye-level on the HMD and power it from a 9V battery (Fig. 1). The use of a HMD with a micro camera mounted on it prefigures video see-through Augmented Reality applications.

![Fig. 1. HMD with tracker and micro-camera.](image)

2.1 Augmented Reality

Augmented Reality (AR) environments combine real images with computer generated images, making possible the visualization of data, models and meta-information [7]. The AR setting used in this work is the most common one: video see-through, which uses a conventional HMD to render virtual elements on the images captured by a camera.

Our experiment consists in showing data from their own field of expertise to different users, and obtaining their feedback. Two different configurations are used: in the first one, called AR (Augmented Reality), the user sees, through
the HMD, virtual data aligned with a real marker (a black square with known geometry, which can be easily moved by
the user, as shown in figure 6) captured by the camera. In the second configuration, AD (Augmented Desktop), the same
augmented data are shown in a desktop display and the view is updated according to the position and orientation of a
hand-held video camera the user can move around the markers (Figure 7).

Our Augmented Reality prototype uses the ARToolkit [8], a graphical toolkit capable of detecting markers in a scene and
render virtual objects according to the orientation and position of the markers. The orientation of the camera is extracted
directly from the detected markers. Since ARToolkit only allows rendering of VRML models, we use the Visualization
Toolkit (VTK) [9] for the processing and visualization of different types of 3D data. This toolkit is very useful since it
offers several functionalities for the manipulation and visualization of 3D data, import of models and data, and
interaction with devices. Basically, we used ARToolkit only to extract the camera pose from the video images and we
render data and models correctly according to the markers using VTK. This configuration provides more flexibility and
allows testing several different data sets and visualization methods with minimal effort.

2.2 Virtual Reality

The VR test environment is an immersive game, which consists of corridors defining a maze in which the player has to
navigate and catch 21 objects in a pre-defined period of time (5 minutes). This environment uses a HMD for display,
mouse buttons and orientation sensors for interaction. Users play the same game in a traditional desktop environment
(VD), using mouse and keyboard as interaction devices and the desktop display.

Most of the programming was done in C++ using OGRE [10], a well-known open-source rendering engine. Some
additional code was written to perform collision detection and counting/recording of events within the game. The same
game is played in both Desktop (VD) and Virtual (VR) environment but with a different interaction. The mouse
(VD)/tracker (VR) is used for the orientation of the head (where the player is looking at), whereas keys (VD)/mouse
buttons (VR) are used for navigating inside the maze (where the player goes).

3. AUGMENTED REALITY EXPERIMENT

3.1 Experiment

The first experiment in Augmented Reality consisted in visualizing scientific data. Two data sets were used, the first
corresponds to the water flow around a ship hull: the pressure at each point is mapped through colour, and arrows are
used to represent the water velocity around the hull (see Fig. 2). The second data set corresponds to the temperature
values inside an oven, represented as coloured points (see Fig. 3).

![Fig. 2. The water-flow data set.](image)

Two other experiments were performed with medical data. In the first data set, the users are interested in locating
epileptic foci by studying the position and orientation of electrical activity within a human brain and use a dipole
representation (vectors located within the patient’s head) as shown in Fig. 4. The last data set corresponds to human
lungs containing air bubbles (see Fig. 5). In this example, lungs are represented as transparent volumetric information to
allow the visualization of bubble anomalies inside the lungs, and users are mainly interested in the relative volume of
these anomalies.
For each data set, users were provided with different viewpoints and scaling factors to make visualization easier. Data were visualized in the Augmented Reality (AR) environment as well as in the Augmented Desktop (AD), in order to compare the differences between the two environments and assess which one provides the users with better manipulation, orientation and visualization experiences. We also provided users with the possibility to visualize the background camera image with different opacities (from totally visible to totally black, in which case only the data set is shown).

In the AR experiment the user holds the marker and sees the data lying on the marker and can interact by moving his head, as well as the marker, to change the viewpoint (Figure 6). While in AD mode, the user holds a camera in his hand and moves it around the marker to change the viewpoint while observing the updated image on the monitor (Figure 7).
In these experiments, we used a qualitative evaluation: users were asked to use the application and then express their opinions by talking to the evaluator while, or after, they were performing the tasks. The evaluator could comment and counterargument, but at the end users stated their overall preferences. We had the collaboration of four domain experts and all experiments started by the AR environment and then proceeded to AD.

Fig. 6. The AR experiment with a clone copy of the image seen by the user on the monitor. Fig. 7. The AD experiment: image captured by the hand-held camera and projected on the monitor.

3.2 Results

First of all, we were surprised with the positive response of the participants to the Augmented Reality (AR) environment, perhaps due to their enthusiasm in using a new and attractive form of visualization. All users showed great interest in the application, suggested improvements and proposed future work with other kinds of data sets.

Concerning manipulation and orientation issues, all users concluded that the AR environment with HMD was more intuitive than the Augmented Desktop (AD). This was due to the facts that users had some difficulties to synchronize the movements of the hand-held camera with the markers in AD, whereas in AR the micro-camera is fixed on the HMD and both hands were used to manipulate the markers. On the other hand, AD was found better than AR regarding the quality of visualization, due to better resolution and contrast, a result that we already expected.

To conclude the experiments we asked the users suggestions on how to improve the system. The more relevant were:

1. The possibility to navigate inside the data in a totally immersive virtual environment;
2. A pointer to indicate specific data features to other users;
3. An option/button that would allow the user to zoom in and out of the data;
4. The possibility to display a thin square representing the border of the marker, to help identify its location without disturbing the viewing of the data (although users did not get lost when the displayed background is black and the camera image used for tracking is not shown).

The overall conclusion was that, for short periods of time, as when somebody wants to obtain a general idea of the data displayed, or to show something to a user that is not used to that kind of data, the Augmented Reality environment makes data manipulation easier since the interaction is more intuitive. However, for long periods of work the Augmented Desktop environment could be preferable since the display has a higher resolution and its use is less tiring.

As a consequence of the users suggestions, two additional experimental conditions were implemented. One adds a virtual thin white line around the marker, when the background camera image is not shown. The other shows the camera image instead of the black background, when the user loses sight of the marker. Finally, we also added stereoscopic vision.

These modifications were presented to the same users. The results were that the thin square does help to manipulate the data without losing sight of the marker, but when it disappears from the camera field of view, seeing the real image helps to find the marker faster. So the users preferred to combine the two suggested modifications.
Regarding stereoscopic vision, the noticed improvement depends on the data set. According to the users, it was not significant in the water flow data set (probably because it was a dense vector field), but in the medical examples (where a volume is presented, either brain or lungs) there was a significant improvement: users got a better sense of depth and thus could better understand the data. The distance between the user’s eyes was found a relevant parameter, as one of the users had great difficulties to sense depth until this distance was set to a smaller value [3].

4. VIRTUAL REALITY EXPERIMENT

4.1 Experiment

As mentioned, the game developed for this experiment can be played in two different modes. In VR mode, the user’s head movements and the mouse buttons are used to control the walking direction (right button forward, left button backward), each user being allowed to hold the mouse in his/her dominant hand (see Figure 8). The desktop version (VD) was played in a way similar to most point and shoot games: the mouse was used to select the moving direction, and two keyboard keys were used to go forward or backward.

During the game the user had to navigate in a virtual maze having the goal to locate and collect as much flying boxes that appeared on the player’s path as possible (see Figure 9). The maze complexity was significant enough to avoid the creation of a mental map. This was done on purpose to avoid, as much as possible, learning from one experience to the other. In practice, most players moved around in a more or less random way during the 5 minutes the experience lasted.

This study was simultaneously qualitative and quantitative. Users started by using the application in both VR and VD versions and, afterwards, they were asked a few questions about their profile (as age, profession and game playing habits, experience in using 3D), as well as their opinion and preferences regarding the two playing modes. During the interview, additional questions were asked such as the degree of disturbance caused by the cables, if the preference for VR was due to novelty, which kind of games users prefer, suggestions of improvements.

We started from a general simple hypothesis that the performance and satisfaction of users was similar in both environments and used a within-groups experimental design [11], i.e., all subjects perform under both experimental conditions, VR and VD. This experimental design has the advantages of requiring less subjects as well as reducing the effect of individual differences. To avoid a possible bias on the results due to learning or boredom effects, the order in which the conditions are tackled was varied between users; thus users were randomly divided into two groups; one started by using the VR environment and the other started by the VD. During the game, in both environments, several parameters (dependent variables) were recorded: total gaming time, number of collected objects, number of collisions with the walls, walked distance, and average speed.
Forty-two volunteer users (twelve females and thirty males) aged from fourteen to forty (the majority was between sixteen and thirty) played the game in both environments. The experiment started with a short introduction to the game, its goals, the used hardware components (HMD basically) and a reference to the questionnaire to be completed at the end of the experiment. It was also pointed out to the users the fact that they could abandon the experiment at any time. After this initial presentation, the user started to play the game without any previous training. During the experiment, an observer was monitoring the user’s performance and taking down relevant comments (such as difficulties in catching flying objects).

4.2 Results

In this section we present the main results obtained from the analysis of the data collected during the experiment through Exploratory Data Analysis, non-parametric tests and Multivariate Analysis, all performed using STATISTICA [12].

The first variables analysed were the number of caught objects, number of collisions, distance and velocity. After a preliminary Exploratory Data Analysis (EDA) using box-plots [13], we decided to test the equality of medians by means of non-parametric tests since none of the data have normal distribution. Figure 10 shows, on the left, the box-plot corresponding to the number of objects caught by all users and, on the right, the number of collisions. The median values of the number of caught objects are 14.5 for VR and 17 for VD. A Wilcoxon Matched Pairs Test [14] rejected the hypothesis of equality of medians (p=0.000002). This means that the difference between the number of objects caught by the users in both environments is statistically significant and cannot be due to chance. On the contrary, the median values of the number of collisions (31.5 for VR and 31 for VD), shown on the right of figure 10, are not significantly different since a test of the same type didn’t reject the hypothesis of median equality (p=0.880730).

The box-plots corresponding to distance and average speed are shown in figure 11. Also in these cases, Wilcoxon Tests rejected the hypothesis of equality of median values of distance (543.5 for VR and 819.5 units for VD; p<0.000000) and speed (6.6 units for VR and 9.9 for VD; p<0.000000).

![Fig. 10. Number of objects caught by the users (left), and number of collisions (right).](image1)

![Fig. 11. Distance walked by the users (left), and average speed (right).](image2)
From these results we can conclude that users had globally better performances when using the VD than when using the VR environment, i.e., they caught more objects, attained higher speeds and walked larger distances, with roughly the same amount of collisions. This general result was already anticipated.

In spite of the fact that the main independent variable of this experiment was the test environment (with two different levels: VR or VD), some secondary variables that could have been influencing the obtained results were identified from the onset of the experiment; thus, we decided to investigate their influence on the obtained results. These variables are related to the users’ profile (as gender, and experience in computer games, and viewing images of 3D data, models or scenes), or to working conditions (as learning effects from one environment to the other). Since the correlation between walked distance and average speed in both environments is very high (1 for VR and 0.96 for VD), we considered only one of them. Moreover, as the number of collisions is not significantly different in both environments, we studied the number of objects caught and the average speed categorized by gaming experience (Figure 12), experience with 3D (Figure 13) and gender (Figure 14).

In order to study the influence of those variables, users were categorized according to their experience, taking into account their answers to the questionnaire and making groups of users, as balanced as possible. Thus, in spite of having asked users how they classified their experience in computer games and 3D (not experienced, moderately experienced, and highly experienced), we have categorized experience in computer games using only two levels (not experienced - 0, experienced - 1) and with 3D using the three levels (not - 0, moderately - 1 and highly experienced - 2). This categorization produced groups having a number of users varying from 11 to 18 which, for the statistical analysis purposes, seems reasonable.

Figure 12 shows the number of caught objects and the average speed for the two user groups according to their experience in computer games. We can observe that both variables have higher values for the VD (hatched boxes) than for the VR environment, for all groups of users, as it was already found for all users in figure 10 and 11. Analysing the results of the users with and without experience, we notice that neither the objects caught nor the speed attained vary much between the groups for VR. However, the same is not true for the desktop (VD), in which case users having experience caught significantly more objects (p=0.01) and attained higher speeds (p=0.04), than non-experienced users.
Figure 13 shows the number of caught objects and the average speed for the three user groups according to their experience with 3D. As in the previous case, we can observe, that both variables have higher values for the VD than for the VR environment, for all groups of users, as it was already found for all users in figure 10 and 11. The number of objects caught by the users in both environments (VR and VD) was not influenced by this categorization. On the other hand, the equality of average speed among the three groups was rejected (Kruskal-Wallis: p=0.03 for VR and p=0.004 for VD). These results could mean that previous experience with 3D was relevant in speed to users in both environments but experience in computer games was relevant only when using the VD environment.

Regarding gender, figure 14 shows the number of objects caught and average speed attained by males and females. The box-plots suggested, and it was statistically supported (Mann-Whitney Test) that both groups seem to have similar performances in the VR environment, but, on the contrary, males outperformed females in the VD. A notice of caution is due to the fact that the two groups are not balanced (twelve females and thirty males) and males have more experience in computer games, which can make invalid the apparent result that females could perform better in the VR environment.
In order to alleviate learning effects between the two environments, half of the users (22) started the experiment by the Virtual Desktop and the others started by the Virtual Reality environment. Nevertheless, we tried to investigate if there was an increase on the attained speed or on the number of objects caught on the second time the users were playing the game, comparing their performances with the performances of users playing the game on the same environment but at a first time. This comparison was performed analysing the box-plots of figures 15 and 16, and performing Mann-Whitney tests in order to assess if the differences between medians observed in the box-plots are statistically significant. These figures show the number of objects caught (Figure 15) and the speed attained (Figure 16) by the users in each environment categorized by the environment where they started the experiment (i.e., in which they have played the game for the first time). The results obtained by the tests could imply that there is a learning effect only on the speed: users playing the game for the second time seem to attain a higher speed than the ones playing in the same environment for the first time. This learning effect seems to exist in both directions (from the VD to the VR environments and vice-versa). This result seems to contradict the idea, presented in [15], that users only transfer experience from the VR to the VD environment; this could be explained by the different conditions in both experiments (such as the use of a 6DOF tracker, and the visualization always in the HMD).

A few other relevant aspects were asked the users after the experiment: if they had felt disoriented or sick, and their degree of satisfaction with the experience in both environments. The results were that more users felt disoriented and sick in the VR environment (19 and 25 users respectively) as compared to the VD environment (5 and 10 users); however, the degree of satisfaction was comparable (31 users enjoyed very much both experiences).

Finally, in informal conversation afterwards, many users mentioned that the cables connecting the hardware they were wearing in the VR environment (glasses and mouse) were uncomfortable, hence suggesting we use wireless devices. Some users reinforced this suggestion explaining that they had felt embarrassed to play the game using the VR environment in public, mainly because the cables were often entangled around them as they turned while exploring the maze.

5. CONCLUSIONS

The supporting technologies of Virtual and Augmented Reality environments still face great challenges. Image contrast and resolution are still much better in a desktop display, which is also less tiring compared with the HMD; in addition, hardware cables also increase the lack of comfort.

Despite these limitations, most of the users in the Augmented Reality experiment preferred the HMD setting, since it offers them a more intuitive and easy interaction when compared with the desktop set-up. Users considered the AR set-up particularly suitable for short periods of data analysis or for data presentation to other users outside the field, who won’t need to learn how to interact with specific hardware or software, since the AR environment is intuitive.

Regarding the VR experiment, we can conclude that global performance was better for the traditional desktop set-up. This result might be affected by the fact that subjects were using a standard configuration for game interaction known to them, and a VR environment that most of them had never used before. However, some performance results were comparable (for instance, in terms of median, the main goal, catching objects, just differs from 14.5 to 17). A few users, mainly those who seldom play computer games, even performed better in VR mode (caught more objects).

Whereas the Augmented Reality user study was application-specific, since it involved users analysing their own data, we consider the Virtual Reality experiment as an abstract user study. Thus, how the results of the Virtual Reality experiment transfer to a real application domain is not obvious, without a formalism that can relate the abstract tasks evaluated to the specific real-world application. Another limitation of this study is the fact that the results obtained from categorization (according to several secondary variables) might be less relevant than other results, given that some of the groups have a relatively small number of subjects.

These experiments showed that the natural interaction offered by and AR/VR system can bring gain to the users, despite the limitations of the technology. With better hardware, the gains will probably be even larger, when compared with the traditional desktop environment. Still, each environment has its advantages and limitations, making each one of them more suitable for different applications.
6. ACKNOWLEDGEMENTS

The authors wish to thank Filipe Teixeira Dias, Alexandre Cruz, José Maria Fernandes and José Silvestre Silva for their time and contribution to the Augmented Reality study, as well as all the users that participated in the Virtual Reality study.

The stay at the University of Aveiro of Frank van Huussen, Jan-Willem Baggerman and Pauline van der Horst was done within the framework of the Erasmus Programme of the European Union.

REFERENCES