Usability evaluation of non-immersive, desktop, photo-realistic virtual environments

Rochelle Villanueva[1], Antoni Moore[2] & B.L. William Wong[3]

[1]Office of the Dean, School of Business
University of Otago. Dunedin, New Zealand
Phone: +64 21 442-528 Fax: +64 3 479-8171
Email: rvillanueva@business.otago.ac.nz

[2]Spatial Information Research Centre
University of Otago. Dunedin, New Zealand
Phone: +64 3 479-8138 Fax: +64 3 479-8311
Email: amoore@infoscience.otago.ac.nz

[3]Interaction Design Centre, School of Computing Science
Middlesex University, London, United Kingdom
Phone: +44 (0)20 8411 2684 Fax: +44 (0)20 8411 5216
Email: w.wong@mdx.ac.uk

Presented at SIRC 2004 – The 16th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
November 29th-30th 2004

ABSTRACT

This study is about the evaluation of non-immersive desktop web-based photo-realistic virtual environments using think-aloud protocol and heuristic evaluation to determine three aims: (1) whether applying the same usability evaluation methodologies result in additional usability guideline categoriesidentified from the Koykka, Ollikainen, Ranta-aho, Milszus, Wasserroth and Friedrich 1999 study; (2) whether think-aloud protocol or usability heuristic evaluation is a better evaluation method for identifying usability problems in desktop, photo-realistic virtual environments; and (3) whether large-scale desktop, photo-realistic virtual environments will have more usability problems than small-scale non-immersive, desktop, photo-realistic virtual environments. The results show that using the Emergent Theme Analysis (ETA), four broad themes were derived: functionality, interaction, appearance and user comments with functionality and interaction broad themes being similar to the categories suggested by Koykka *et al.* (1999). Furthermore, the results indicate that small-scale non-immersive, desktop, photo-realistic virtual environments had more usability problems than large-scale non-immersive, desktop, photo-realistic virtual environments, contrary to the hypothesis that large-scale VEs will have more usability problems due to its complexity and that the think-aloud protocol derived more themes compared to the heuristic evaluation – suggesting that TAP is a better usability evaluation method than HE in this type of study. However, a combination of the two qualitative methods has identified a greater number of usability problems, supporting the need for triangulation of research methods. The investigation resulted in new design guidelines that will allow for more usable design of non-immersive desktop, photo-realistic virtual environments. Furthermore, the study provides some new areas for future developments of usability evaluation methods for non-immersive desktop, photo-realistic virtual environments.

***Keywords and phrases*:** usability evaluation, virtual environments, heuristic evaluations, think-aloud protocol, photo-realistic virtual environments

1.0 INTRODUCTION

The concept of Virtual Reality (VR), a three-dimensional, computer-generated environment that allows for single or multiple users to interact, navigate, respond, and experience a synthesized world modelled from the real world, has provided social, scientific, economic and technological change since its inception in the early 1960's. Since that time, VR has also evolved into many forms and taken different tangents. One form that it has taken is non-immersive desktop photo-realistic Virtual Environments (VEs), where real-world still images are joined together to create 360 degree panoramas of places and objects in time. The concept has become increasingly popular since its creation using Apple QuickTime Virtual Reality Authoring Studio (QTVRAS) in 1995. Although there have been many evaluations conducted on the design and usability of 3-Dimensional (3D) VEs for various applications e.g. (Darken & Sibert 1996; Bowman & Hodges 1997; Johnson 1998; Dalgarno & Scott 1999; Koykka, Ollikainen, Ranta-aho et al. 1999; Vinson 1999; Chen 2000; Sutcliffe & Deol Kaur 2000; Stanney, Mollaghasemi, Reeves et al. 2003), there seem to be limited guidelines available in the design of non-immersive, desktop photo-realistic VEs.

This study is about the evaluation of non-immersive desktop web-based photo-realistic virtual environments using a previously applied evaluation method for non-immersive desktop 3D multi-user environments conducted by Koykka, M. *et al*., (1999). The two qualitative evaluation methods are think-aloud protocol and heuristic evaluation. Think-aloud protocol is a usability test where the participants are asked to verbalize their thoughts, feelings, and opinions while interacting with the system (Axup 2002; Zhang 2003). While a heuristic evaluation is an informal usability inspection method for finding usability problems in a user interface by having a small set of evaluators examine an interface and judge its compliance with usability principles (heuristics) (Molich & Nielsen 1990; Nielsen 1994).

There are three aims in this study:

- to determine whether applying the same usability evaluation methodologies result in additional usability guideline categories: 3D environments should provide support for orientation, navigation and movement, real world metaphors need to be clearly understandable and avoidance of delay and waiting periods in performance (identified from the Koykka *et al.* study);
- to determine whether think-aloud protocol or usability heuristic evaluation is a better evaluation method for identifying usability problems in desktop, photo-realistic virtual environments; and
- to determine whether large-scale non-immersive, desktop, photo-realistic virtual environments (i.e. of a large building, such as Sydney Opera House) will have more usability problems than small-scale non-immersive, desktop, photo-realistic virtual environments.

2.0 THE STUDY

Twelve subjects took part in evaluating two different web-based photo-realistic VEs (small- versus large-scale) using the Think-Aloud Protocol (TAP) and Jakob Nielsen's usability Heuristic Evaluation (HE). A variation of the grounded theory research method called Emergent Themes Analysis (ETA) was used to "distil" the recorded narratives into broad themes.

2.1 The Virtual Environment (VE) systems

The virtual environment systems used in this experiment were from the World Wide Web (WWW). A basic set of criteria was applied in determining which system to choose. These are as follows:

- The virtual environment was web-based for easy accessibility,
- The virtual environment was photo-realistic and viewable in at least 180 degree angle.
- The VE systems had similar scale and complexities such as the number of location levels
- The VE system had similar number of nodes available within each system
- The VE system had similar presentation styles such as the use of interaction mechanisms such as navigational aids were similar.

All systems allowed the use of interactive maps and embedded hotspots with which to navigate. All systems also presented additional information relevant to a user's location in the VE system. The Sydney Opera House and Questacon Science Center had a similar number of nodes (40 or more locations). While Wroxton Abbey and Richard Strauss House had similar number of nodes (less than ten locations). Thus they formed natural groupings for the design study.

Table 1 shows the four web-based VE systems, their complexities and presentation styles (the shaded areas represent large-scale VEs, while the non-shaded areas represent small-scale VEs) .

Table 1.Web-based Photo-realistic VE systems evaluated in this research.

Name and Location	Web address	Number of Floor Levels	Number of Nodes (Locations)	Interaction mechanisms present (Navigational Aids such as maps, embedded arrows, pop-up labels etc)
Sydney Opera House Sydney, Australia	http://www.soh.nsw.gov.au/virtual_tour/vrtour.html	4	approximately 82 nodes	3D map with links, embedded hotspots with hotspot indicators, orientation arrow on the 3D map, automatic VE rotation without mouse holding.
Questacon Science Center Canberra, Australia	http://virtual.questacon.edu.au	5	approximately 40 nodes	3D map with links, embedded hotspots with hotspot indicators, orientation arrow on the 3D map, automatic VE rotation without mouse holding.
Wroxton Abbey Oxfordshire, England	http://www.wroxtonabbey.org/index1.html	1	approximately 7 nodes	2D map with links, embedded hotspots with labels on the control bar
Richard Strauss' House Garmisch, Bavaria	http://www.richardstrauss.at/html_e/02_privat/0fs_index.html	1	approximately 7 nodes	3D map with links, embedded hotspots, with labels on the control bar

Furthermore, Figures 1 and 2 show example screenshots a large- versus small-scale VE system.

Figure 1. Large-scale VE System - the Sydney Opera House .

Figure 4. Small-scale VE system - the Richard Strauss House.

2.2 Research Methods

2.2.1 Think-Aloud Protocol (TAP)

Think-Aloud Protocols (or verbal protocols) are often used when trying to identify how people use an interface or toolset. Participants are asked to speak aloud as they work through some task. Each choice used, button pressed, etc is verbalised, as well as what the participant is looking at during the process. All sessions are usually video recorded and analysed later. This technique has both positive and negative aspects. It is a good technique to use when trying to ascertain if there is a particular problem within an interface or when there is a process change and measurement of learning time is required. However, there may be unexpected results primarily because people are not used to slowing down their thinking to allow for speech (Axup 2002; Zhang 2003).

2.2.2 Heuristic Evaluation (HE)

Heuristic evaluation is an informal usability inspection method of finding usability problems in a user interface by having a small set of expert usability evaluators examine an interface and judge its compliance against usability principles also known as heuristics. The technique was developed by Jakob Nielsen and Rolf Molich (Nielsen 1997) in which experts guided by a set of usability principles known as heuristics, evaluate whether user interface elements of an invented telephone index system conform to usability and design principles (Molich & Nielsen 1990; Nielsen 1994).

It is important to note that the order in which the evaluation was conducted is critical. A think-aloud protocol was performed before a heuristic evaluation. Performing the latter evaluation method first would affect the participant's familiarity and learning of the system. For example, any initial difficulties with using interaction techniques such as navigation would have already been discovered through initial exploration and identification using the usability heuristics. This may result in the participant anticipating and compensating any difficulties when using the system during the think-aloud protocol. As this research is determining which evaluation method would identify a greater number of usability problems, initial issues may not be deemed important by the participant and therefore, will not be voiced out.

2.2.3 Emergent Themes Analysis (ETA)

Emergent Themes Analysis (ETA) is "an iterative distillation process" developed by Wong and Blandford (2002) to help extract design insights by identifying themes and decision strategies from voluminous interview data from real-time operational environments (namely emergency ambulance dispatch). The ETA approach is based on Grounded Theory but tailored to take advantage of the exploratory and efficient data collection features of the Critical Decision Method (CDM), a retrospective interview technique requiring participants to recall a memorable incident that they have been involved in (Klein, G., Calderwood, R. et al. 1989), (Wong, B. L. W. 2004) when investigating the nature of decision making in the complex socio-technical domain of emergency ambulance dispatch (Wong, B. L. W. and Blandford, A. 2002). Figure 3 outlines the key steps taken in an ETA approach.

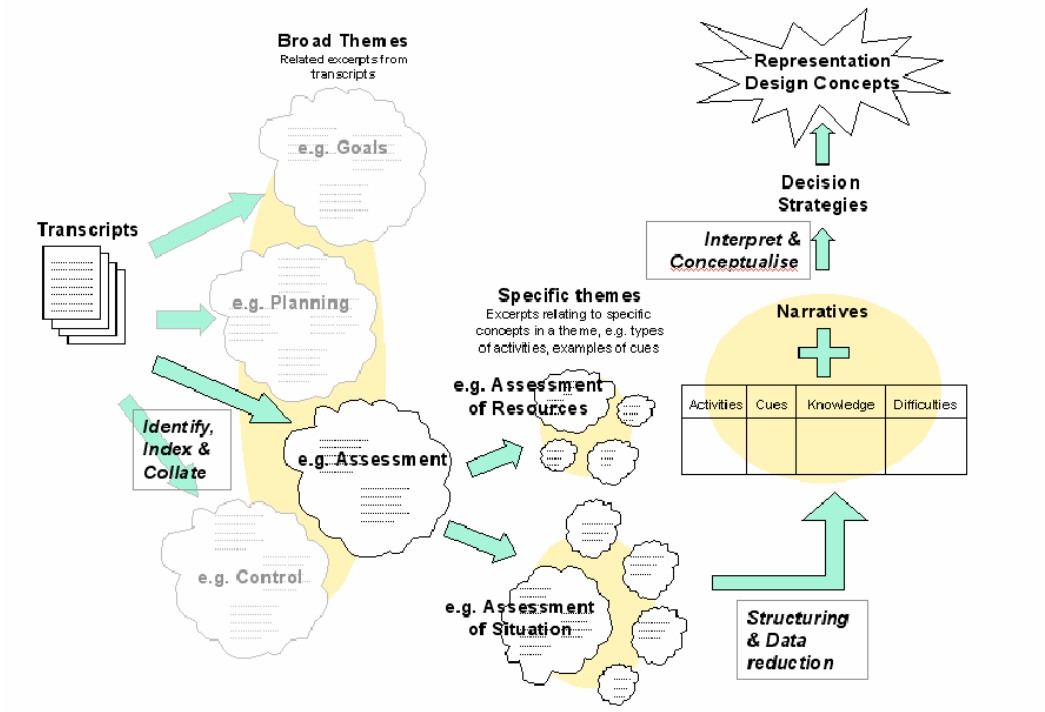

Figure 3. Emergent Themes Analysis process

Broad themes are similar ideas and concepts reported across interviews and incidents. They are identified, indexed and collated. In this way, themes emerge from the interview data. In Wong and Blandford (2002), broad themes were initially associated with broad aspects of the decision process, such as the goals that an ambulance allocator needs to achieve. Using a similar distillation procedure, sub-themes, or specific themes within each

broad grouping were identified, and the data and the specific themes are further categorized according to a framework for describing each decision process in finer detail. This framework has four categories that describe activities, cues, knowledge and difficulties of the processes, as illustrated in Figure 6-3. The specific themes and supporting data are then distilled into summary tables and narratives, which are descriptive prose that consolidated the data even further. The narratives and the summary tables are then interpreted, and relationships and new ideas about decision strategies are conceptualised. These decision strategies are again in narrative form, and written so as to provide understanding about the needs for representation.

1.0 ANALYSIS AND RESULTS

1.1 Data Analysis with Emergent Themes Analysis Approach

1. Each individual participant transcripts is reviewed looking for user difficulties encountered when using the VE.
2. Once reviewed, each transcript will be analyzed to identify common remarks made by the users. The following narratives below can be classified under the issue "Lack of Feedback" (where P## is the participant number and (0:00:00) is the time stamp). For example:

> P01 (0:06:53) - So when I press on it, suddenly the thing (popup text) comes up and sometimes it did not come up automatically and I have to wait a little while. So in the first place I'm trying to figure out what are these dots.
>
> P02 (0:15:43) - Okay, actually to see this map, what does it (rolling over the 3D map) okay, I think this tell me about my position inside the house (rolling over the 3D map). No, No, it's not my position. OK, What actually is in this position? I'm not sure where I can find my position here, I trying to find out.
>
> P10 (0:59:23) - (Rolling over the different icons on the VR window). Umm, I'm wondering what is this thing (rolling over the link for high resolution panoramas), it's an icon for a real player. Click on it to see what it does (click on the QuickTime icon). Okay, it's loading an image, but I have to figure out how to use it now, just by possibly clicking (clicked on the image, no effect). Nothing is happening, and I'm wondering what I was supposed to do (0:59:59)...
>
> P11 (0:01:18) - "And I'm not quite sure where I am (panning within the VR and as you pan, yellow arrows and "i" for further information appears inside the VR indicating that it is click-able)."

3. Once all similar difficulties are gathered from individual participants, they are collated and re-analyzed across all the participants. Similar concepts are then collated. The collation below, illustrate how narratives are grouped together.

> The response time for activating something is a little slow during which time there is nothing to indicate to the user what is happening (Feedback) in this case what its the yellow balls for a different location
>
> P01 (0:06:53) - So when I press on it, suddenly the thing (popup text) comes up and sometimes it did not come up automatically and I have to wait a little while. So in the first place I'm trying to figure out what are these dots.
>
> P10 (0:59:23) - (Rolling over the different icons on the VR window). Umm, I'm wondering what is this thing (rolling over the link for high resolution panoramas), it's an icon for a real player. Click on it to see what it does (click on the QuickTime icon). Okay, it's loading an image, but I have to figure out how to use it now, just by possibly clicking (clicked on the image, no effect). Nothing is happening, and I'm wondering what I was supposed to do (0:59:59)...
>
> User is not sure of their current position/ location
>
> P02 (0:15:43) - Okay, actually to see this map, what does it (rolling over the 3D map) okay, I think this tell me about my position inside the house (rolling over the 3D map). No, No, it's not my position. OK, What actually is in this position? I'm not sure where I can find my position here, I trying to find out.

> P11 (0:01:18) - "And I'm not quite sure where I am (panning within the VR and as you pan, yellow arrows and "i" for further information appears inside the VR indicating that it is click-able)."

Following the above initial groupings, each collection of similar narratives is conceptualized into a problem. For example:

> The user does not know what the VE system is doing. There is a lack of feedback to the user especially when loading.
>
> The user does not know what they did to achieve the current system state
>
> The user does not know their current position – spatial location, orientation

4. Finally, all the identified problems are then reviewed and analyzed for a common theme. In this case, the broad theme identified was "Lack of VE System Status Feedback."
5. Steps 1-4 are repeated again with the remaining data until there are no more broad themes that can be identified.

3.2 Results

The results show that ETA derived four broad themes derived from using the ETA method. The following section also presents some of the sub-broad themes that emerged to create the broad themes:

- Functionality - limited or lack of VE system status feedback, hidden system functionality, hidden interaction object functionality, limited freedom and movement within the VE system, lack of cross-platform availability, lack of distinct functions between novice and expert users and help and documentation;
- Interaction - user expectations of real world conventions were not met, lack of system exploration, limited use of the provided navigational aids, user difficulty in navigating and panning within the VE and determining their current location, looking for useful information and desired locations, lack in clarity between the map and the VE relationship, and limited field of view;
- Appearance - information, VE presentations including image quality and location and size of hotspots; and
- User comments - VE comments and suggestions such as to be able to view objects in greater detail, and move through the nodes as if they were walking.

The results also show that the functionality and interaction broad themes were similar to the categories suggested by (Koykka, Ollikainen, Ranta-aho et al. 1999). For example, the functionality sub-broad theme of limited or lack of feedback such as the current status of the system provides support Koykka's "Avoidance of delay and waiting periods in performance". Help and documentation, limited use of navigational aids and hidden system and interaction object functionality and freedom and movement contribute to the difficulty of users in determining their current location, to move, navigate and pan within the virtual environment, and therefore interact with the VE, supporting Koykka, M. *et al.*, support for "providing support for orientation, navigation and movement". In addition, the absence of real world conventions (expected objects and behaviors) within the VE also posed difficulty in navigating and panning within the VE and the users' limited FoV also supports Koykka, M. *et al*., finding that "Real world metaphors need to be clearly understandable".

Also, the results indicate that small-scale non-immersive, desktop, photo-realistic virtual environments had more usability problems than large-scale non-immersive, desktop, photo-realistic virtual environments, contrary to the hypothesis that large-scale VEs will have more usability problems due to its complexity.

Using the ETA method to analyze the data, the think-aloud protocol derived more themes compared to the heuristic evaluation – suggesting that TAP is a better usability evaluation method than HE in this type of study. However, a combination of the two qualitative methods has identified a greater number of usability problems, supporting the need for triangulation of research methods.

4.0 CONCLUSION

The investigation resulted in new design guidelines that will allow for more usable design of non-immersive desktop, photo-realistic virtual environments. Furthermore, the study provides some new areas for future developments of usability evaluation methods for non-immersive desktop, photo-realistic virtual environments. Regardless of its future use, usability evaluations of *any* virtual environments are still needed.

REFERENCES

Axup, J. 2002, *Comparison of Usability Evaluation Methods (UEM)*, Published on-line at Date Accessed UserDesign.com.

Bowman, D. A. 2002, 'Principles for the design of performance-oriented interaction techniques', .

Bowman, D. A. & L. F. Hodges (1997) An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. *1997 Symposium on Interactive 3D Graphics*, April 1997 pp. 35 - ff.

Bowman, D. A., D. B. Johnson & L. F. Hodges (1999) Evaluation of Virtual Environment Interaction Techniques. *VRST'99: ACM Symposium on Virtual Reality Software and Technology*, University College, London, December 20-22, 1999 pp. 26-33.

Chen, J.-L. (2000) A model of wayfinding in virtual environments: Strategies for navigational aids. *Dissertation Abstracts International: Section B: the Sciences & Engineering*, 60:11-B, pp. 5701, US: Univ Microfilms International.

Dalgarno, B. & J. Scott (1999) Usability Problems in Desktop Virtual Environments. *OZCHI'99 Interfaces for the Global Community*, Charles Sturt, Wagga Wagga, Australia, pp. 134 - 136.

Darken, R. P. & J. L. Sibert (1996) Wayfinding Strategies and Behaviours in Large Virtual Environments. *Computer-Human Interaction (CHI)*, .

Goldbaum, H. 2003, *The Virtual Wroxton Abbey*, Published on-line at Date Accessed Oxfordshire, England, Wroxton College.

Johnson, C. 1998, *On the Problems of Validating DesktopVR Innovative User Interfaces: Multimedia and Multi-modal User Interfaces, Wearable Computers and Virtual Reality*, Published on-line at Date Accessed Proceedings of the HCI'98 Conference on People and Computers XIII.

Koykka, M., R. Ollikainen, M. Ranta-aho, W. Milszus, S. Wasserroth & M. Friedrich (1999) Usability Heuristic Guidelines for 3D Multi-User Worlds. *Interfaces for the Global Community*, Charles Sturt University, Wagga Wagga, Australia, pp. 52 - 57.

Molich, R. & J. Nielsen (1990) Improving a Human-Computer Dialogue. *Communications of ACM*, 33:3, pp. 338- 348.

Nielsen, J. (1994) Usability Inspection Methods. *CHI'94*, Boston, Massachusetts, USA, April 24-28, 1994 pp. 413-414.

Nielsen, J. 1997, 'Usability Testing', in *Handbook of Human Factors and Ergonomics*, Ed G. Salvendy, John Wiley and Sons, pp. 1543-1568.

RichardStrauss 2003, *The Villa in Garmisch*, Published on-line at Date Accessed Garmisch, Bavaria, Richar Strauus.

Sandra, L. 2002, *Task Analysis*, pp 1-6.

Saretto, C. J. (1997) A Survey of Interaction Issues Facing Navigation in Virtual Worlds. *CHI 97: Conference on Human Factors in Computing Systems*, Los Angeles, USA, April 18-23, 1997 pp. 1-7.

Stanney, K. M., M. Mollaghasemi, L. Reeves, R. Breaux & D. A. Graeber (2003) Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. *International Journal of Human-Computer Studies*, 58:4, pp. 447-481.

Sutcliffe, A. & K. Deol Kaur 2000, *Evaluating the Usability of Virtual Reality User Interfaces*, Published on-line at Date Accessed Behaviour and Information Technology.

Vinson, N. G. (1999) Design Guidelines for Landmark to Support Navigation in Virtual Environments. *CHI'99: The CHI is the Limit ACM SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA USA, May 15 - 20, 1999 pp. 278-285.

Zhang, Z. W. 2003, *Usability Evaluation Methods*, Published on-line at Date Accessed Drexel University.