Case Study: Sequential Evaluation of the Virtual Prints Concept and Pilot Implementation

A. Mourouzis[1], D. Grammenos[1], M. Filou[1], P. Papadakos[1], C. Stephanidis[1, 2]

[1] Institute of Computer Science, Foundation for Research and Technology-Hellas, Science & Technology Park of Crete, Heraklion, Crete, Greece

{mourouzi, gramenos, papadako, filou, cs}@ics.forth.gr

[2] Department of Computer Science, University of Crete, Heraklion, Crete, Greece

Abstract

Virtual Prints are a mechanism for supporting navigation, orientation, way-finding and a number of other functions in virtual environments. This paper presents the results of an evaluation experiment of a prototype virtual environment equipped with part of the Virtual Prints mechanism. The objectives of the study were to assess the prototype's usability, and to further elaborate on the concept and the potential functionality offered by Virtual Prints. The evaluation was carried out using both immersive and non-immersive systems, and was based on a sequence of evaluation methods including expert-based review, cooperative evaluation (i.e., including both experts and users), and user-based tests. The paper briefly introduces the reader to the concept of Virtual Prints and presents the employed approach and the tools used. Then, an overview of the findings of the study, both qualitative and quantitative, is provided along with the related emerging challenges for the future. The paper concludes with a critical review of the approach that was followed, providing insights on the methods and tools used, as well as on the study results and their impact on the improvement of Virtual Prints..

1. Introduction - Background

The concept of *Virtual Prints* (ViPs) along with a corresponding software mechanism have been introduced in (Grammenos, Filou, Papadakos & Stephanidis, 2002) as a handy tool for navigation support (travelling, way-finding, orientation, feedback and history) in Virtual Environments (VEs). ViPs can be considered as the digital counterparts of real-life tracks that people leave behind in their environment. Three different manifestations of ViPs have been suggested (Grammenos et al., 2002):

- ***Virtual Footprints*** *(ViFops)*: they are left behind while the 'inhabitants' of a virtual world are moving in it;
- ***Virtual Fingerprints*** *(ViFips)*: they are 'imprinted' on VE objects and structural elements every time someone interacts with them;
- ***Virtual Fossils*** *(ViFossils)*: special marks that can be permanently left within the virtual space, or on any object, upon user request and can be considered as a kind of personal landmarks.

In addition to the above, ViPs can provide, in an intuitive and user-friendly way, several functions and concepts that are popular, if not standard, in conventional 2D applications, such as: interaction shortcuts, bookmarks, help, interaction and navigation history, undo / redo and repeat, annotation, highlighting content, and marking / identifying (non) visited virtual places (Mourouzis, Grammenos, Filou, Papadakos & Stephanidis, 2003).

This paper presents the evaluation of a prototype implementation of the ViPs Mechanism[1] that was conducted on the one hand, to assess its usability and, on the other hand, to further study the concept, in terms of intuitiveness, required and potential functionality. The evaluation was carried out using both an immersive and a non-immersive version of the system and comprised a sequence of expert-based review, co-operative (i.e., involving both experts and users) and user-based experiments. The paper presents the evaluation methods and tools used for the assessment of this Prototype, and provides an overview of the findings of the study, both qualitative and quantitative. Emerging challenges for the future are also presented. The paper concludes with a critical discussion of the evaluation approach and of the problems encountered, providing valuable insights on the methods and tools used, and highlighting the impact of the results on the concept and implementation of Virtual Prints.

2. Overview of the ViPs Prototype

The ViPs Prototype (Figure 1) is a single user system; nevertheless it simulates a collaborative virtual environment as the user 'shares' the virtual world with a number of artificial VE 'inhabitants' (human avatars controlled by the system, moving around randomly). The presented "world" used is quite simplistic and consists mainly of a single entrance/exit maze along with some simple interactive objects, such as doors, buttons and three-dimensional geometrical shapes. All the 'inhabitants' of this world, including the user, leave behind their ViPs as they move around and interact with the VE. ViPs are generated and recorded by a ViPs Mechanism integrated in the developed pilot system. Two different computer setups were used for the needs of this study: (a) an immersive VR system using a stereo HMD (Virtual Research V8, with 640x480 true VGA LCDs); and (b) a desktop system using a 17" monitor. Both versions were running on Dual Pentium III 1Ghz PCs with Linux Slackware 8.0 with a Geforce-Ti4200 graphics card, using a conventional 2D mouse with three-buttons for navigation and interaction.

In the developed Prototype, the user interface of the ViPs Mechanism has been instantiated through 3D menus, buttons, and interactive consoles (i.e., panels). Each ViP that is released is associated to an *Information Sheet* (Grammenos et al., 2002) that can be viewed when the mouse cursor is over it. The *Information Sheet* includes information about the type, the owner, and the creation time of the ViP. When a ViP is selected (using the right mouse button, since the left is used for travelling), if it does not belong to the 'current' user, a menu appears offering options such as: viewing the *ViPs Navigation Console* (Mourouzis et al., 2003); viewing the *User Scale/Hide Console* (Mourouzis et al. 2003); and modifying ViPs configuration parameters (e.g., display options). In brief, the *ViPs Navigation Console* allows the user to track the ViP's owner and navigate within the virtual space using the existing ViPs (e.g., follow their path), while the *Scale/Hide Console* allows the user to modify the ViPs display configuration, in general or for a specific group of ViPs (e.g., those belonging to a specific user). In case the ViP belongs to the 'current' user, then by selecting it, in addition to the aforementioned options, the user can change the ViP's creation options. In addition to the above, a menu is always accessible through a 3D button residing at the bottom of the user's viewpoint and allows the user to: (a) release a new ViFoP or ViFossil; (b) activate / deactivate the *ViPs Mechanism*; (c) modify the configuration of ViPs; (d) view the *Scale/Hide Console*.

[1] An improved and updated version, based on the original prototype.

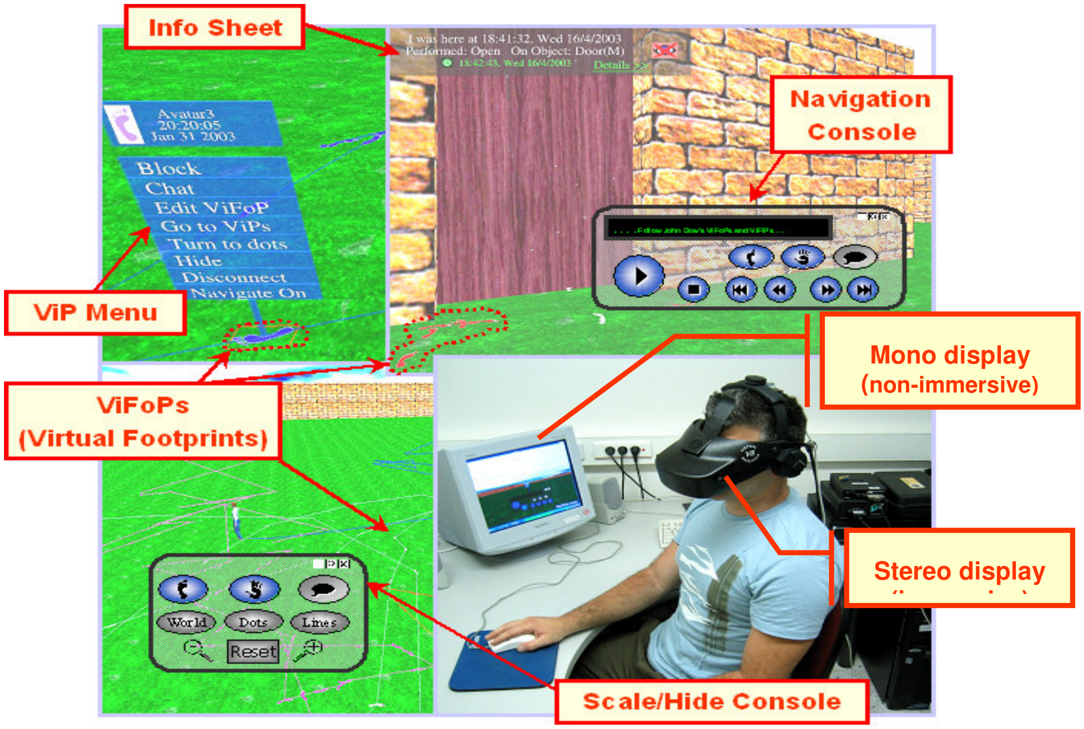

Figure 1: Overview of the ViPs Prototype

3. Evaluation Procedure: Methods & Tools

Virtual Environments require a more rigorous evaluation process than conventional 2D applications. The distinctive differences of VEs with respect to conventional 2D user interfaces give rise to a number of issues that need to be addressed. Bowman, Gabbard & Hix (2001) provide a thorough overview of the distinctive issues of VE evaluation and of the employed evaluation methods worldwide. Most of these methods have actually been developed for the usability evaluation of conventional applications and have been adapted to support VE evaluation. The methods selected to carry out the evaluation of the ViPs Prototype follow a *Sequential Approach* (Gabbard, Hix & Swan, 1999). As implied by its name, this approach involves a series of evaluation techniques that run in sequence (see Figure 2), such as *Cognitive Walkthrough* (Wharton, Rieman, Lewis, & Polson, 1994), *Heuristic Evaluation* (Nielsen, 1994), formative and summative evaluation (Scriven, 1967; Hix & Hartson, 1993). In general, a sequential evaluation uses both experts and users, produces both qualitative and quantitative results, and is application-centric.

In the following sub-sections, each of the individual methods employed for the purpose of the sequential evaluation of ViPs is briefly described, highlighting how they were adapted for the needs of the study. All the experiments presented were conducted in the *Human Computer Interaction Laboratory* (HCI Lab) of the *Institute of Computer Science* at the *Foundation of Research and Technology-Hellas* (ICS-FORTH) in Crete, Greece (www.ics.forth.gr/hci/).

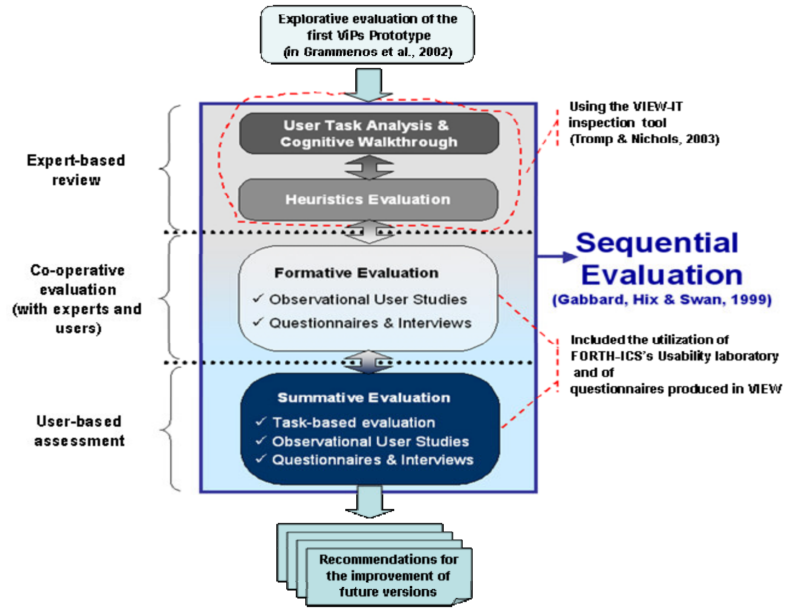

Figure 2: Overview of the sequential approach followed for the evaluation of the ViPs Prototype

3.1. Expert-based Evaluation

The evaluation procedure started with an expert-based inspection of the utility and usability of the ViPs Prototype. In this step, *VIEW-IT* (Tromp & Nichols, 2003), a rapid usefulness assessment tool designed to assess the usefulness of VEs in terms of utility and usability, was used. The evaluation team consisted of five assessors (3 male, 2 female) with a rich background in usability engineering and interaction design of 2D applications. All of them are members of the HCI Lab of ICS-FORTH. The assessors were first given a specific list of simple tasks to drive a *Cognitive Walkthrough* (Wharton, Rieman, Lewis, & Polson, 1994) and then they were allowed to use freely either of the two versions of the system (immersive and non-immersive) in order to evaluate the interface and judge its compliance to utility heuristics and general usability principles (*Heuristic Evaluation* (Nielsen, 1994)). All assessors worked concurrently, and at some stages co-operatively. The whole procedure lasted around 5 hours, including breaks. The results emerging from this inspection were also used to prioritise development issues towards bringing the system to an appropriate level for running user-involved experiments (see Figure 3 and sections 3.2 and 3.3).

3.2. Co-operative Evaluation

Following the expert-based evaluation, the ViPs Prototype was further refined and a *Co-operative Evaluation* (Wright & Monk, 1992) was conducted. In this evaluation step, twenty people (8 female, 12 male) were asked to try out the proposed interface using a desktop version of the system by freely exploring it and performing some simple tasks. The *Co-operative Evaluation* is actually an extension of the 'think-aloud' verbal protocol, where users are encouraged, in addition to think-aloud, to ask any questions regarding the system, the assessed functionality, or the tasks that they are asked to perform during the test. The two evaluators who conducted this evaluation step were free to ask the participants questions at any time during the tests. In our study, assessor – participant discussions and user (re-) actions and comments were recorded using digital audio and video, as well as by the means of debriefing interviews and pre- and post-hoc questionnaires aimed at collecting more specific information on the participants and their experiences with the system. A booklet composed of a number of questionnaires and interview forms was created and distributed to all the participants. This booklet also included the forms used for the user-based evaluation (see the next section). For the needs of this test, the booklet included: (a) a pre-test questionnaire for collecting background information about the participants (i.e., a demographics and experience questionnaire); (b) a post-test questionnaire for assessing the usability of the ViPs pilot based on a *Usability Questionnaire* for VEs developed by the Virtual Reality Applications

Research Team at the University of Nottingham (see Patel and Nichols paper in these proceedings for further details).

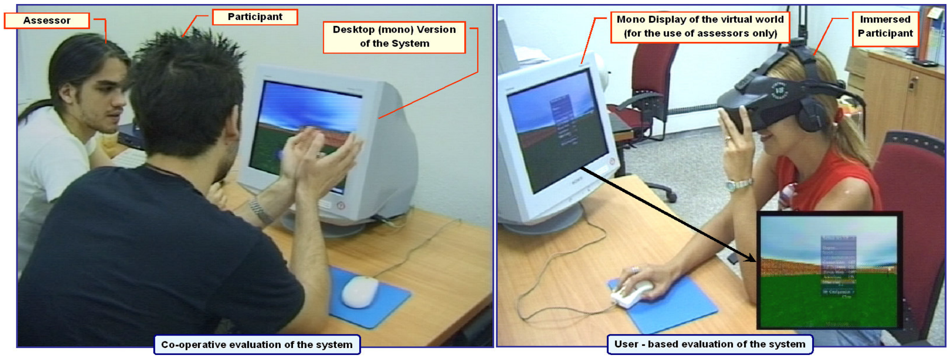

Figure 3: Tests involving users

An additional aim of this experiment was to give the participants enough time with the desktop version of the system to learn the basics of the ViPs Mechanism (i.e., learning by exploration) prior to using the HMD-based version of the system (see section 3.3). Finally, aiming at transforming the produced qualitative data into quantitative data, a slight variation of the method suggested by Marsh and Wright (Marsh & Wright 1999) was employed, in which empirical evaluations were assessed on the basis of the quantity and quality of the users' 'think-aloud' verbalizations.

3.3. User-based Evaluation

All twenty people involved in the co-operative evaluations were asked to participate in a task-based usability evaluation of the ViPs Prototype (within a maximum period of four days since the first evaluation). This final evaluation step aimed at: (a) identifying usability problems and assessing the ViPs ability to support user exploration, learning, and task performance; (b) identifying alternative usage patterns of the ViPs Mechanism. Both formative and summative evaluation methods were used. The users had to perform specific task scenarios, while user errors, actions in the VE, physical reactions and comments were recorded by means of digital audio, a digital camera, and a video converter device. For the needs of this experiment, the participant's booklet mentioned above included: (a) pre-test questionnaires consisting of an *Immersive Tendency Questionnaire* (Witmer & Singer, 1998), a *Stress Arousal Checklist* (Gotts and Cox, 1988) and a *Simulator Sickness Questionnaire* (Kennedy et al, 1993); (b) a *Short Symptom Checklist* (Nichols et al, 2000a)to be filled in every five minutes during the test; and (c) post-test questionnaires consisting of a *Presence Questionnaire* (Witmer & Singer, 1998), the same *Stress Arousal Checklist*, *Simulator Sickness Questionnaire*, and *Short Symptom Checklist* as in the pre-test questionnaires (for comparative studies), a similar *Usability Questionnaire* to the one mentioned in section 3.2, an *Enjoyment Questionnaire* (Nichols et al, 1999), and a *Post-immersion Assessment of Experience Questionnaire* (Nichols et al, 2000b).

The significant differences between the co-operative evaluation and the user-based evaluation of the ViPs Prototype were that, for the latter, the participants: (a) were given specific tasks to perform; and (b) used the HMD version of the system. The average immersion time (using the HMD) per user was around 30 minutes with a five minute break after the first quarter. Thus, a total of 20 hours of video were produced for analysis (20 participants, 2 half-hour videos for each one – a video capturing the user interacting with the system and the video converter capturing the user interacting within the virtual world).

4. Results Overview and Discussions

4.1. Expert-based review results

The evaluation team used the VIEW-IT usefulness inspection tool (see section 3.1) to assess the utility of the system in terms of its practical use, as well as its usability. VIEW-IT defines three aspects of utility for VR/CAD hybrids: "a good degree of presence, a good degree of user-friendliness and a low degree of VR-induced sickness" (Tromp & Nichols, 2003). In this study, the reported level of presence experienced whilst using the ViPs Prototype was found to be moderate. It was suggested that this could be improved by: (i) increasing the level of detail and the quality of the display (e.g., using an HMD of higher resolution); (ii) by using visual and auditory feedback; and (iii) by increasing and stabilising the system response rate. On the other hand, the level of *VR-induced sickness* was low. According to the assessors, this could possibly be lowered even more by keeping active user movement to a minimum (i.e., minimise the need for movement whilst interacting with the ViPs interface). Finally, the overall *usability* (learnability, efficiency, memorability, errors, user satisfaction - Nielsen, 1993) of the system was also found to be moderate, and a number of suggestions were produced for improving it, including: (a) providing adequate and consistent feedback, as well as cues to permissible interactions; (b) introducing error messages and error management; and (c) improving the quality of graphics. Overall, the assessors concluded that the functionality and support provided by the tested *ViPs Mechanism* can increase the usability of a VE, both in terms of ease of learning and user satisfaction, but also in terms of efficiency, memorability, and user error rate.

4.2. Co-operative evaluation results

Several results emanating from the co-operative evaluation confirmed the existence of certain design issues that had already been identified during the expert-based evaluation. A considerable number of conclusions were derived from the analysis of the post-evaluation questionnaires used in this step. However, the majority of the results that were collected, emerged from the study of the videos captured (assessor-participant discussions, etc.). Table 1 provides an overview of the number of recorded and analyzed verbalizations per participant, a statistical analysis of which (based on average values) is presented in Figure 4.

<table>
<tr><th></th><th>Study Time (mins)</th><th colspan="2">Overall verbalisations</th><th colspan="2">Usability-related verbalisations</th></tr>
<tr><td rowspan="7">Average (Median) values for the 20 participants sample</td><td rowspan="7">33:09
(34:12)</td><td>Total Number</td><td>106,11 (90)</td><td>Total Number</td><td>11,05 (12)</td></tr>
<tr><td rowspan="2">A. Think aloud verbalisations</td><td rowspan="2">54,74 (45)</td><td>1. Learnability</td><td>3,03 (3)</td></tr>
<tr><td>2. Efficiency</td><td>3,50 (4)</td></tr>
<tr><td rowspan="2">B. User to Assessor Question</td><td rowspan="2">25,58 (22)</td><td>3. Memorability</td><td>1,03 (1)</td></tr>
<tr><td>4. Errors</td><td>0,00 (0)</td></tr>
<tr><td rowspan="2">C. Assessor to User Question</td><td rowspan="2">25,79 (25)</td><td>5. User Satisfaction</td><td>3,50 (3)</td></tr>
<tr><td>% over total verbalisations</td><td>10,4% (12,8%)</td></tr>
</table>

Table 1: Overview of the average (and median) verbalizations made per participant during the co-operative evaluation

In general, most participants agreed that ViPs are easy to learn and use and can be a handy tool for navigation, way-finding and annotation. All participants enjoyed using the ViPs mechanism. The participants could not clearly identify the degree to which ViPs can support orientation, although the analysis of the videos showed that ViPs actually provided considerable help regarding user orientation. Other interesting figures in this study include the total study time (about 11 hours of video were analysed); the total number of recorded verbalizations (more than 2100) and the total number of usability-related verbalizations identified (around 220 of which about 100 were identified by both assessors as having high value for the design and use of ViPs). It is also interesting to note (left part of Figure 4) that if only the think-aloud method was employed, then about half of the collected information would have been missed (since think-aloud produced only 51,6 % of the total verbalisations).

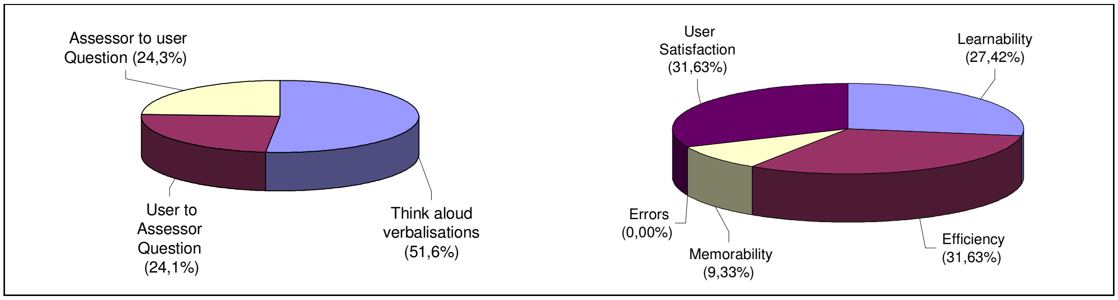

Figure 4: Statistical overview of the overall (left) and usability-related (right) verbalizations made during the co-operative evaluation (based on the average values presented in Table 1).

		Evaluator 1		
		LOW	HIGH	TOTAL
Evaluator 2	LOW	116	10	126
	HIGH	10	92	102
	TOTAL	126	102	228

Table 2: Confusion Matrix

Further to the method suggested by Marsh and Wright (in Marsh & Wright 1999), the two evaluators involved in this evaluation step were asked to judge the quality (high or low) of the identified usability problems to designers of ViPs. A confusion matrix showing the level of confidence on the quality judgements is shown in Table 2. Using Cohen's Kappa (K), the results of the confusion matrix provide a significance rating of 82%, which according to this formula gives an 'excellent' degree of inter-rate reliability for the result (Robson, 1995).

4.3. User-based evaluation results

This last evaluation step reinforced the study's pre-evaluation hypotheses, and the results of formative evaluation in line with the findings of the two previous steps. For instance, the study of the *Usability Questionnaire* mentioned above, had, among others, the following results (e.g., which comply in a satisfactory level with those for learnability, user satisfaction and efficiency presented in Figure 4):

- **Hypothesis: The implemented *ViPs Mechanism* is easy to learn**. The question on the ease of learning scored 73,75 %, (0% means that the users totally disagree with the statement, while 100% that they strongly agree).
- **Hypothesis: The implemented ViPs mechanism increases the subjective user satisfaction**. The question concerning subjective user satisfaction (related to using the ViPs concepts and mechanism) scored 72,5%.
- **Hypothesis: The implemented ViPs mechanism is generally a handy tool for VEs**. The related question scored 70%.

Almost all the participants reported that they found the scenarios they had to perform easy to learn and remember, and still motivating and entertaining. On the other hand, the technology used (e.g., the HMD, the mouse, the workstations, etc.) constitutes a serious negative factor to the usability of the ViPs pilot. Several comments were received regarding the quality of the graphics (related to the HMD screen resolution), the difficulties related to the mouse as a navigation device and to system response. Although the focus of this study was not directed towards such issues, these parameters had to be taken under consideration during the statistical analysis. Overall, the results of this final step did not produce any conflicts with the findings of the two previous steps, and also provided complementary input allowing us to obtain a holistic view of the overall usability of the ViPs Prototype.

5. Conclusions

In general, the sequential approach followed in this study, and the evaluation tools used to support it, have proved to be very effective and efficient in studying the concept of ViPs and their pilot implementation. Some suggestions for further improvement of the overall evaluation procedure followed include:

a) The use of a test for assessing subjective spatial abilities, such as the Guilford-Zimmerman Test (Guilford & Zimmerman, 1948) to assess whether participants who score higher on the spatial orientation subset of the Guilford-Zimmerman Test outperform those with lower scores.

b) Since the prototype was evaluated with a rather limited number of users (20), mainly with high computer expertise, it would be beneficial to test its intuitiveness and usability with a more wide and diverse (in terms of age, computer experience and background) sample of participants.

c) The use of head tracking for the user-based (i.e., HMD-based) tests. It is expected that this would significantly improve the usability of the system allowing us to focus more on ViPs-related issues.

Regarding ViPs, the results from all experiments (expert-based, co-operative, and user-based) have facilitated: (a) the refinement of the concept; (b) the enrichment of the ViPs Mechanism behaviour and functionality; (c) the collection of suggestions for improvement of the user interface. Since the experiments conducted had complementary objectives and outcomes, the evaluation team would not have been able to arrive at the same results by using just a single method. Overall, the sequential approach to usability evaluation of VEs is highly recommended for exhaustive assessments both of research prototypes and industrial VE applications.

In conclusion, the study presented in this paper is part of an ongoing effort towards: (a) the refinement and improvement of a novel interaction concept (i.e., ViPs) for VEs and its digital manifestation; (b) the establishment of a structured process for achieving this; and (c) contributing to the evolving research domain concerned with the evaluation VEs. In light of the findings of this study, it is strongly believed that a fully-functional ViPs Mechanism can significantly increase the usability of VEs. In this direction, future work shall seek to develop the envisaged ViPs Mechanism and integrate it to existing VE systems from diverse application domains and assess its impact on their usability.

6. Acknowledgments

Part of the work reported in this paper has been carried out in the framework of the European Commission funded *project IST-2000-26089-VIEW of the Future* ("Virtual and Interactive Environments for Workplaces of the Future").

7. References

Bowman, D. A., Gabbard, J., & Hix, D. (2001). Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. Technical Report TR-01-17, Computer Science, Virginia Tech.

Gabbard, J.L., Hix, D., & Swan E.J. (1999). User Centered Design and Evaluation of Virtual Environments, IEEE Computer Graphics and Applications 19(6), 51-59.

Gotts, G. and Cox, T. (1988) Stress and arousal checklist: A manual for its administration, scoring and interpretation. Swinburne Press.

Grammenos, G., Filou, M., Papadakos, P., & Stephanidis, C., (2002). Virtual Prints: Leaving trails in Virtual Environments. In Proc. of the Eighth Eurographics Workshop on Virtual Environments, 30-31 May 2002, Barcelona, Spain.

Guilford, J.P., & Zimmerman, W.S. (1948). The Guilford-Zimmerman Aptitude Survey. In Journal of Applied Psychology, 32, pp. 24-34.

Hix, D., & Hartson, H.R. (1993). Developing User Interfaces: Ensuring Usability through Product & Process, John Willey and Sons, New York.

Kennedy, RS., Lane, NE., Berbaum, KS. and Lilienthal, MG. (1993) Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. The International Journal of Aviation Psychology 3(3), pp 203-220.

Marsh, T., & Wright, P. (1999). Co-operative Evaluation of a Desktop Virtual Reality System, Workshop on User-Centred Design and Implementation of Virtual Environments, King' s Manor, University of York, pp. 99-108.

Mourouzis, A., Grammenos, D., Filou, M., Papadakos, P., & Stephanidis, C. (2003). Virtual prints: An Empowering tool for Virtual Environments. In the Proc. of the Human Computer Interaction International Conference 2003 (HCII2003), 22-27 June 2003, Crete, Greece.

Nichols, S., (1999). Virtual Reality Induced Symptoms and Effects (VRISE): Methodological and Theoretical Issues. PhD Thesis: University of Nottingham.

Nichols, SC, Haldane, C. and Wilson, JR. (2000a) Measurement of Presence and its consequences in Virtual Environments. International Journal of Human-Computer Studies (2000), 52, pp471-491

Nichols, SC., Ramsey, AD, Cobb, S., Neale, H., D'Cruz, M. and Wilson, JR. (2000b) Incidence of Virtual Reality Induced Symptoms and Effects (VRISE) in desktop and projection screen display systems. HSE contract research report 274/2000.

Nielsen, J. (1993). Usability Engineering. Academic Press, Inc., San Diego, pp. 26-37.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen and R. Mack (Eds.) Usability inspection methods. Wiley, New York.

Robson, C. (1995) Real World Research: A resource for Social Scientists and Practitioners – Researchers, Blackwell, Oxford.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), Perspectives of curriculum evaluation, American Educational Research Association Monograph, Rand McNally, Chicago.

Tromp, J., & Nichols, S. (2003). VIEW-IT: A VR/CAD Inspection Tool for use in Industry. In Proc. of the Human Computer Interaction International Conference 2003 (HCII2003), 22-27 June 2003, Crete, Greece.

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen and R. Mack (Eds.) Usability inspection methods. Wiley, New York.

Witmer, B.G., & Singer, M.J. (1998). Measuring Presence in Virtual Environments: A presence Questionnaire. IN presence: Teleoperators and Virtual Environments, 7(3), pp. 225-240.

Wright, P. & Monk, A. (1992) Co-operative Evaluation: The York Manual, CHI92 Tutorial, CHI Conference 1992.