Composition of HCI Evaluation Methods for Hybrid Virtual Environments

Marcus F. C. Alencar
Tecgraf / Dept. of Informatics
PUC-Rio, Brazil
malencar@tecgraf.puc-rio.br

Alberto B. Raposo
Tecgraf / Dept. of Informatics
PUC-Rio, Brazil
abraposo@tecgraf.puc-rio.br

Simone D. J. Barbosa
Dept. of Informatics
PUC-Rio, Brazil
simone@inf.puc-rio.br

ABSTRACT
Research about interaction design and evaluation of virtual environment (VE) applications is recent and lacks well-established methods and techniques. The focus of this study is to show the results of applying a composition of methods to evaluate the communicability and usability of the HybridDesk, which provides three environments for the interaction with VEs, supporting 3D scene visualization and navigation, as well as 3D object manipulation and annotation. This study performs a qualitative evaluation of the HybridDesk by applying traditional usability evaluation methods, like heuristic evaluation, usage observation sessions, questionnaires and interviews, as well as the communicability evaluation method (CEM), which is based on semiotic engineering. It then compares the evaluation results of these various methods, demonstrating that they all contribute in distinct ways to the evaluation of a hybrid environment. These results also highlight the importance of compatibility among the various signification systems, produced by distinct designers, which a user needs to interpret and understand during interaction.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, input devices and strategies*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism – *virtual reality*.

General Terms
Human Factors, Experimentation.

Keywords
evaluation, usability, communicability, 3D interaction.

1. INTRODUCTION
Interaction with 3D VEs, although rapidly growing, still lacks critical mass and is distant from standardization. There is still not enough knowledge about the best 3D interaction techniques. In this regard, the definition of adequate evaluation methods poses various challenges, particularly due to the diversity of interaction possibilities with VEs, including new interaction devices, like touchscreens, wands, 3D glasses, HMD (Head-Mounted Displays), etc [12].

Hybrid VEs, which combine 2D and 3D interaction, bring added complexity due to additional interaction possibilities and the transition between environments. Typically, usability evaluation methods were conceived to evaluate 2D interfaces within the traditional desktop (WIMP).

Therefore, there is a need for a method, or a composition of methods, to evaluate the user-system interaction in hybrid VEs. This need is illustrated by a study using the HybridDesk as the target system, equipment that includes three interaction environments: one VE for navigation within the 3D scene, one VE for object manipulation, and the third is a WIMP desktop, for annotation reading and writing.

Considering the need to capture the different and unique aspects of user interaction with the HybridDesk, this study applied a composition of HCI evaluation methods, both usability methods, and the communicability evaluation method, based on semiotic engineering [4]. This composition highlights distinct contibutions of each evaluation method and provides evidences about issues arising from different signification systems used by system designers.

The paper is structured as follows: the next section presents the HybridDesk system. The third and fourth sections present the evaluation methodology and preparation. The fifth and sixth sections present the evaluation results and discussion. The paper concludes with considerations on the evaluation of hybrid systems based on our findings.

2. HYBRIDDESK
The HybridDesk was designed as a low-cost, easy-to-use system to support 3D annotation task in an oil platform visualization application. The technological setup to satisfy these requirements provides a semi-immersive workplace with stereoscopy, headtracking and a wand (a Wiimote tracked by an optical tracker), plus a WIMP interface for creating and reading the annotations.

The three integrated interaction environments of the HybridDesk [3] illustrated in Figures 1 and 2 are the following:

- VR-Nav (MiniCave for 3D Navigation and Selection): a semi-immersive environment that utilizes projection screens and anaglyphic stereo glasses with headtracking for visualization. The wand is the input device used for scene navigation control and 3D object pointing and selection. A

SAC'11, March 21-25, 2011, TaiChung, Taiwan.

3D representation of the wand is created in the 3D scene;

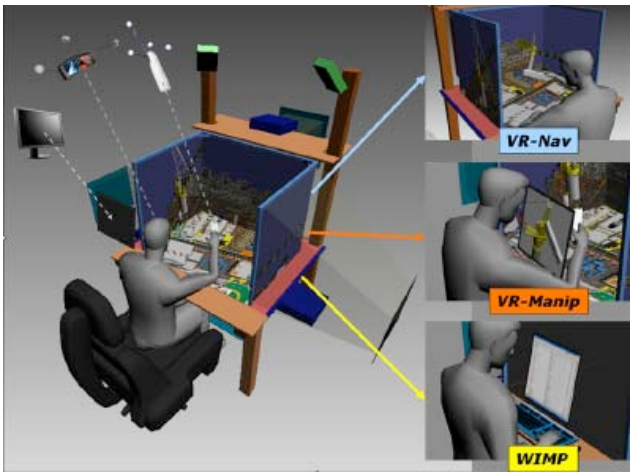

Figure 1. HybridDesk interactive environments.

- VR-Manip (LCD for 3D object manipulation): environment based on the concept of reach-in displays [10]. An LCD was chosen to provide a detailed visual inspection of 3D objects at a shorter distance, like in the real world. The 3D object is visualized at the LCD monitor using stereo glasses with headtracking. The wand is used for object manipulation and to attach an annotation to the selected object;
- WIMP: classical desktop environment. Visualization is available at the four projection screens, plus the LCD monitor. Input devices are the keyboard and mouse.

In order to physically merge these technological setups, a wooden table was built composed of four legs but without the lid (Figure 1). A piece of wood was set in front of the whole structure to provide support for devices such as mouse and keyboard in addition to support for the arms. The description and discussion of these issues were presented elsewhere [3].

(a) (b)

Figure 2. (a) VR-Nav: Navigation Environment (b) VR-Manip: Manipulation Environment [3].

3. THE EVALUATION METHODOLOGY

The composition of evaluation methods for this study was based on the sequence proposed by Gabbard et al. [8] for usability evaluation of VEs, combined with the communicability evaluation method (CEM) [4, 5, 14] plus questionnaires and interviews, resulting in the proposed following activities:

- Expert Guidelines-based (Heuristic) Evaluation [7]
- User Tasks Identification
- Task-based Usage Scenarios
- User Observation with Think-aloud protocol for CEM and Usability [4, 5, 6, 14]
- User Satisfaction and Usability Factors Questionnaire [9, 11, 15, 16]
- User Interview for CEM and Usability [9, 11, 16]

The application of additional evaluation methods tends to increase the overall cost of the evaluation. To keep this impact to a minimum, the approach was to choose methods that could have their execution steps combined as much as possible, particularly the evaluation preparation steps and the user observation execution for both usability and communicability evaluations. This way, it was possible to simultaneously collect data for both evaluations, in common user observation sessions, with users executing the same interaction tasks, without compromising the results produced by each method. Due to the qualitative focus of this evaluation, quantitative evaluation methods were not considered for this study.

At the Evaluation Preparation phase, the following activities were jointly performed:

- Participants Profile Definition and Selection
- Usage Scenario Definition
- Analysis of Ethical Issues
- Preparation of Material for User Observation Session
- Configuration and Adjustments of the Environment
- Pilot Test Execution

At the User Observation phase, the following activities were jointly performed: reception and accommodation of participants; execution of the usage scenario interaction tasks; and post-use interview.

The usability questionnaire was applied right after the usage session, and just before the interview.

4. EVALUATION PREPARATION

4.1 Heuristic Evaluation Planning

The Heuristic Evaluation was performed based on the guidelines of the VE usability taxonomy and suggestions proposed by Gabbard [7], which are organized in four areas: VE Users and User Tasks, VE User Interface Mechanisms, The Virtual Model and VE User Interface Presentation Components. Each of these main areas is divided in sub-areas. However, not every sub-area is applicable to all systems, and therefore not all of the more than 180 suggestions apply as guidelines for every VE system, as even Gabbard has recognized. Therefore, the suggestions that dealt with interface elements that did not exist at the HybridDesk were disregarded, leaving 126 guidelines for consideration by the expert evaluator.

4.2 User Tasks Identification

This step was performed through consultation to the HybridDesk documentation and discussions with the Designer. The user tasks objective was to leave annotations in a 3D scene regarding the modeling quality of an object, as per the manager instructions.

4.3 Usage Scenario Definition

Considering that there are no regular HybridDesk users, a typical usage scenario was defined based on discussions with the Designer and the objectives of the evaluation. The following scenario was handed to the study participants:

"You are an experienced user of 3D model graphic visualization, but have never used HybridDesk before. You were hired by an oil company to verify the modeling quality of 3D objects developed by another group within the company. You will use the HybridDesk on a daily basis to perform the verifications and leave annotations, with your assessment, attached to the verified object. Your manager has left instructions regarding your task in the 3D scene, in an annotation file attached to the top of the staircase at the right side of the heliport of the oil platform. You must navigate at the virtual oil platform using the HybridDesk, find the annotation file and follow the instructions left by your manager in this file."

The interaction tasks involved in the scenario were the following[1], in the environment identified in parentheses:

1. Navigate in the 3D scene up to finding the instructions (VR-Nav);
2. Select the 3D icon representing the annotation (VR-Nav);
3. Read the instructions in the annotation file (WIMP);
4. Navigate in the 3D scene up to finding one of the two objects (VR-Nav);
5. Select the first object (VR-Nav);
6. Manipulate the first object to verify the quality of modeling (VR-Manip);
7. Create an annotation with the quality assessment for the first object (WIMP);
8. Manipulate the annotation icon aiming to attach to the first object (VR-Manip);
9. Manipulate the first object (VR-Manip);
10. Return to the navigation environment and search for the second object (VR-Nav);
11. Select the second object (VR-Nav);
12. Manipulate the second object to verify the quality of modeling (VR-Manip);
13. Create an annotation with the quality assessment for the second object (WIMP);
14. Manipulate the annotation icon aiming to attach to the second object (VR-Manip);
15. Manipulate the second object (VR-Manip);
16. Return to the navigation environment (VR-Nav).

4.4 User Observation Planning

Seven participants (P1 to P7) were recruited, 2 female and 5 male. Their ages varied from 23 to 45, and all of them had some type of previous experience with 3D applications (3D gaming, 3D visualization apps, 3D modeling, etc.) but none of them had previous experience with the HybridDesk. Usability observation sessions were performed with all seven participants, two in excess of the recommended minimum of five participants [11]. The communicability evaluation was only performed with four participants (P1, P2, P3 and P7), as three participants is the minimum required [4] by this qualitative method. Considering that the composition of methods is intended to complement qualitative methods while keeping the evaluation cost low, and not to compare the results in quantitative terms, the different number of participants was not an issue.

During the user observation sessions, the following data were collected: evaluator annotations during observation session; evaluator annotations during interviews; questionnaire about the usage scenario; task logging performed by the system; video and audio recordings of the usage sessions; and audio from the communicability interviews.

The questionnaire answered by the participants after each user observation session was organized in 28 assertions, for which the participants should provide a response in 5-point Likert scale [15, 16], where 1 meant totally disagree and 5 totally agree. Considering the significant number of interaction tasks that should be evaluated, as well as the various usability factors, it would not have been possible to assess every combination of interaction tasks and usability factors in a limited number of questions, so the approach was to produce two groups of questions, one focused on the interaction tasks (1 to 13), and the second group (14 to 28) focused on evaluating the usability factors of key interface elements of the HybridDesk, like the ergonomic aspects of the glasses, the wand and the LCD monitor.

4.5 Communicability Evaluation Planning

Virtual environments and games should engage user's interest and should be usable to permit results that are more clear-cut and enjoyable [13]. Therefore one of the important aspects that should be evaluated in a virtual environment is its communicability. The communicability evaluation method (CEM) [5, 14] focuses in the analysis of the reception quality of the metacommunication message from the system designer to the user. CEM is based on the semiotic engineering theory [4], which brings the designer, users and system to the same communicative context. De Souza and Leitão [5] explains that CEM is performed in five phases: Test Preparation, Test Application, Tagging, Interpretation and Generation of the Semiotic Profile. Test Application corresponds to the User Observation. Tagging is the first analytical phase of the method, where the user behavior during interaction is analyzed and interpreted by the evaluator. The designer-to-user communication breakdowns are identified and tagged according to predefined communication expressions as per Table 1.

Table 1. Communicability Evaluation Tags.

Why doesn't it?	Oops!	I can do otherwise
Looks fine to me	Where is it?	Thanks, but no, thanks
I can't do it this way	What now?	Where am I?
What happened?	I give up	What's this?
Help!		

The Interpretation phase consists of analyzing the occurrence of tags during the interaction, searching for clues as to why each breakdown has happened. The semiotic profile goes beyond the identified communication breakdowns and interaction problems of the previous phases, dealing with a more abstract level, the interface language [14].

[1] The detailed list of tasks was not provided to the participants.

5. EVALUATION RESULTS

5.1 Heuristic Evaluation Results

The expert evaluator considered and classified 126 of the guidelines proposed by Gabbard [7] according to its applicability. Table 2 summarizes this classification.

Table 2. Summary of Heuristic Evaluation Results.

Heuristic Evaluation Results	Quantity
Issues Identified	26
System Complies with Guideline	41
Guideline not Applicable	59
Total Number of Guidelines Considered	126

Each of the 26 identified issues was analyzed and organized in sub-areas, as presented in Table 3.

Table 3. Summary of the Identified Issues by Sub-areas.

Sub-area Issues	Quantity	Sub-area Issues	Quantity
VE Users	3	Manipulation	3
VE Tasks	1	Input	4
Navigation	4	HandHeld	1
Selection	6	System Info	4

5.2 Interaction Tagging (Communicability)

Based on the collected data, including the evaluator notes during the observation, the video and audio recordings of the usage sessions, and the interview, the evaluators analyzed the interaction of the three participants[2] and tagged the communication breakdowns. Table 4 presents a summary of the tagging classification, with the number of tags per tag type and participant.

Table 4. Summary of the Identified Issues by Sub-areas.

Tag Type / Participant	P2	P3	P7	Total
Where is it?	1			1
Looks fine to me	1		1	2
I can't do it this way				0
What happened?	3	1		4
Why doesn't it?	2	1	1	4
What's this?	3	4		7
Help!				0
Oops!	3	1	5	9
Where am I?	2	1		3
What now?	4			4
I can do otherwise	1			1
Thanks, but no, thanks				0
I give up				0
Total Breakdowns	**20**	**8**	**7**	**35**

The use of the think-aloud protocol [6] during user observation has helped to detect various communication breakdowns, especially those that had a short duration, and would appear not to have occurred, except for the fact that the participant spontaneously verbalized his/her temporary reluctance to act with "What now?"

Some tags were only identified during the interview, like in one case of the "What happened?" Although the participant did eventually conclude what had happened to the system, he/she complained at the interview that the system did not offer a clear, easily noticeable feedback.

5.3 Data Interpretation (Communicability)

One of the "What now?" tags was caused by a serious misunderstanding by the participant of how to return to the navigation environment. He/she tried several wrong paths until finally got back on track. The system does not offer an option menu or a help system to guide the user.

The great number of occurrences of the "What's this?" tag had to do with misunderstandings of the representations in the 3D scene. The instruction note left by the manager was to "inspect the tip of the crane". But two participants were confused by which object represented the "tip of the crane" in the 3D scene: the tip of the crane tower, the tip of the crane arm, or the tip of the crane hook? In fact, the intention of the task manager was to say the tip of the crane arm, but two of the participants were not so sure that this was the intent, so they verified the other crane "tips". Another problem with the instruction was the fact that it provided the location of the instruction note in the 3D scene as "the manager instruction is at the staircase at the right side of the heliport." The problem here is that as the whole 3D scene was rotated by the user, so right and left changed directions, but the system offered no navigation orientation hint that defined what was left and right, nor north, south, east and west. It was clear that the system needs to offer some kind of orientation to the user. But which designer should be responsible for providing that? Should the designer of the content provide object labels? Should the designer of the application provide an object search function? Or should the task manager write the instructions using a textual language that better relates to the 3D objects in the scene? This line of questioning has led to the findings that the user interacts with the product of more than one designer, a fact that is exacerbated when interacting with Hybrid VEs, due to the diverse interaction devices and representations, and the great number of designers whose products the user is exposed to.

To facilitate comparison of these results with the usability evaluation, the communicability breakdowns were classified in Table 5 according to the following categories of typical HCI issues:

- Execution: problems in the execution of an interaction action, including aspects of locomotion within the 3D navigation environment;
- Navigation: involves execution planning, higher level navigation, like way-finding and the process of object selection, excluding locomotion actions;
- Signification: problems of signification, meanings attributed to interface elements or the task instructions by the user versus the designer;
- Perception: problems of user perception of the resulting state of the system and/or the actions performed by the system.

5.4 Semiotic Profile (Communicability)

The semiotic profile produced the designer-to-user metacommunication message based on what was perceived by the communicability evaluation participants is the following (main findings in boldface):

[2] Observation data from P1 had to be discarded because the system froze during the observation session, totally disrupting the user experience. As a result, only six participants produced usability data and three produced communicability data.

Table 5. Mapping of Tags to Typical Categories of HCI Issues.

Tag Type/HCI Category	Execution	Navigation	Signification	Perception
Where is it?		1		
Looks fine to me		1	1	
I can't do it this way				
What happened?				4
Why doesn't it?	3			1
What's this?			7	
Help!				
Oops!	7		1	1
Where am I?				3
What now?		4		
I can do otherwise		1		
Thanks, but no, thanks				
I give up				
Total	**10**	**7**	**9**	**9**

"According to my interpretation you are a user with knowledge of 3D modeling, **with relevant experience in VE navigation**. I am assuming that **you will interact with the system using your right hand**.

I understand that you want to navigate in 3D scenes in an environment designed for this purpose, where you can also select objects for manipulation and inspection in a specific environment, and you want to use a wand to navigate and manipulate objects. For annotation, you want to use the WIMP desktop. So I have conceived the HybridDesk with three interaction environments, one for navigation and object selection, one for object manipulation and inspection, and one for annotation creation and lecture. With this system you can perform a series of objectives: navigating in 3D VEs, manipulating and inspecting 3D objects, and associating text annotations to 3D objects.

The *Wand* has some buttons that **you will have learn to use with practice, because I have not made available an on-line help system** on how to use them. There are keys to navigate in the 3D scene, **but when you are far from the object, navigation will seem very slow**.

In the navigation environment you can navigate the 3D scene **using the Wand functions that you have learned by using them**. You can **inspect objects in this environment, although I have conceived the manipulation environment to do so**. You may find annotations represented by a 3D icon attached to objects at the 3D scene.

Interacting with the *Wand* will make you discover that **if you press the A key for too long and you are pointing to an object, you will select the object and leave the navigation environment**. In this case, I will notify you with a brief animation at the MiniCave screen, **so you should pay attention to this warning.**

When at the WIMP environment you will be able to return to the navigation or manipulation environments, **but you will have to remember where you came from and use the specific command for each situation.**

In the manipulation environment you will be able to manipulate the object and attach text annotations to it, which will be represented as a 3D icon at the tip of the Wand in the LCD monitor. I will let you know when the Wand touches the object with a vibration. When you release the annotation on the object, **the 3D icon will not be necessarily visible at the manipulation environment**. If you want to **go back to the navigation environment you will have to remember the specific command to do so**."

5.5 Usage Observation Results (Usability)

As Table 6 shows, a large number (24) of usability issues were caused by execution (E) problems, in particular due to wand button manipulation difficulties, wand buttons functionality memorization, especially the use of the Home button to return from manipulation to navigation.

Table 6. Usability Issues during the User Observation Sessions (E=Execution, N=Navigation, S=Significance, P=Perception).

#	Usability Issues Description	E	N	S	P
1	Lost orientation when navigated through the crane model		1		
2	In doubt of which were the left and right sides of the Heliport			1	
3	Went initially to the wrong stair			3	
4	Difficulty to approximate to the plataform when far away	4			
5	Selected the 3D icon correctly but with difficulty	1			
6	Was initially disoriented when returned to the original position		1		
7	Found the glasses uncomfortable	1			
8	Thought that it would be difficult to identify the crane			1	
9	Arm was tired	1			
10	Turned around the whole 3D scene to verify the tip of the crane				4
11	Selected the tip of the tower instead of the tip of the crane			1	
12	Selected the hook thinking it was the tip of the crane			2	
13	Preferred not to move the LCD to the center	2			
14	Pressed the A button unintentionally when disrupted by the LCD (left-handed)	1			
15	Created a new file because could not find the previously created file				1
16	Notepad message was visually obstructed by the LCD				1
17	Difficulty to see the wand at the LCD due to the left-handed position				1
18	Did not notice if the 3D icon was attached to the object or not				1
19	Verified the tip of the crane but did not leave a message when it had no cracks			3	
20	Terminated the session without verifying the second crane		1		
21	Tried to delete the attached annotation but had no success	1			
22	Selected again the annotation icon because had forgotten the task		1		
23	Selected again the annotation icon to verify the task was correctly done				1
24	Selected an object instead of the annotation icon	2			
25	Selected an object unintentionally	2			
26	Was lost, not knowing how to return to navigation		1		
27	Tried to interact with manipulation as if it was in navigation				2
28	Went to WIMP unintentionally, due to a slip in the A button	1			
29	Was in doubt if the annotation was attached or not, and reselected				1
30	Pressed the A button twice and returned to WIMP	1			
31	System was stuck in WIMP because name of the file was changed	1			
32	Went to manipulation unintentionally, due to a slip in the A button	1			
33	Went to WIMP intentionally, but wanted navigation	1			
34	Pressed the A button too long and went straight to WIMP	3			
35	Tried to return using the wand commands				1
36	Tried to return using the back option	1			
37	Returned to WIMP without noticing what was going on				1
	TOTALS	**24**	**5**	**11**	**14**

It is worth noting that problems with the input devices (mouse and keyboard) are not common in typical desktop systems.

One of the issues with greater execution incidence (4 out of 6 people) was the difficulty of traveling at the 3D scene from a distant point of view. This is a scale problem, for which the designer should find a solution. Another difficulty was to make the selection of very small objects, in particular the 3D icon. Any small arm movement deviates the selection to another object. But a large part of the execution problems were ergonomic, like fatigued arms, discomfort with the glasses and LCD manipulation. In the latter case, two participants chose not to slide the LCD towards the desk center for object manipulation, using the LCD at its left-side position. On the other hand, a left-handed participant was forced to move the LCD to the desk center, so that he/she could manipulate the object in the LCD using the wand at his/her left hand, but had some difficulties in moving the LCD.

The second largest number (14) of usability issues was caused by perception (P) problems, indicating the need to improve system feedback. Participants had difficulties to notice the result of their actions, which caused involuntary environment transitions that were not clearly perceived, although the system did provide visual animations as a feedback of transitions. One participant that did not notice the transition to the manipulation environment so he/she continued to interact with the system using the wand, as if he/she was still in the navigation environment. The fact that the navigation screens of the mini-CAVE remained apparently active when in the manipulation environment has misled this participant. The lack of visual feedback when the 3D icon is attached to an object has also caused problems to some participants. The third largest number (11) of usability issues was caused by signification (S), which was surprising, especially considering that in many instances the cause was the interpretation by the user of the meaning of the task, as defined by the task manager, represented by the evaluator in the test.

Three out of six participants went initially to the wrong staircase, because they did not understand the task correctly, which told them to look for the annotation at the "...staircase to the left of the Heliport." This kind of issue raises various questions, particularly regarding which was the real cause of this misinterpretation, and how to avoid it. One difficulty is the need to translate instructions from a textual message to a 3D scene, where the staircase had only a visual representation. The lack of orientation marks or maps in the 3D scene did not help either. The participants also had some difficulty in interpreting what the actual meaning of "tip of the crane" in the instructions was, some confusing it with the tip of the crane tower, and others confusing it with the hook hanging from the crane. The "Home" button at the wand had also different meanings, depending on the context, which also caused confusion to some participants.

These occurrences brought up the evaluator's attention to the fact that the user interaction with the HybridDesk actually involved at least three distinct signification systems: "the system" (wand commanding), "the content" (3D scene) and "the task" (textual instructions). A different person produced each of these significations systems: the system designer, the 3D modeler and the task manager. In the standard desktop applications, this interaction inconsistency is not so visible or critical, because most signification systems rely extensively on textual labels, absent or infrequent in VEs.

The smallest number (5) of interaction issues was attributed to actual navigation (N) problems, where the user knew what to look for but did not know how to find his/her way in the 3D scene. This is why many of the issues identified, although they occurred in the navigation environment (VR-Nav), were classified in other categories.

5.6 Questionnaire and Interview (Usability)

Table 7 presents the answers to questions 1 to 13 of the usability questionnaire. The results show the least satisfaction in selecting the 3D icon, followed by object manipulation. The tasks performed at the desktop (WIMP) produced the highest satisfaction levels, as they were very familiar to the users. The questionnaire responses indicated a high level of user discomfort with the stereoscopic glasses, what could have been aggravated by the fact that most users already used glasses. Interviews also acquired qualitative information to understand the reason for the participants' choices on the usability questionnaire.

Table 7. Answers to the Usability Questionnaire – Tasks.

Usability Questions - Tasks	P2	P3	P4	P5	P6	P7	Avg.
1. It was simple to navigate at the 3D scene	4	3	4	4	4	3	3.67
2. It was simple to select the 3D icon	2	1	3	2	4	3	2.50
3. It was simple to select objects	4	5	3	4	5	4	4.17
4. It was simpe to manipulate objects	3	4	3	4	3	2	3.17
5. It was simple to attach the 3D icon to a object	5	5	2	3	5	4	4.00
6. It was simple to read the file with the description of the task	5	5	4	5	5	4	4.67
7. It was simple to create files with annotations	5	5	4	5	5	4	4.67
8. It was simple to transition from the VR-Navigation environment to the VR-Manipulation environment	5	4	4	3	5	3	4.00
9. It was simple to return from the VR-Manipulation environment to the VR-Navigation environment	2	4	4	2	5	4	3.50
10. It was simple to transition from the VR-Manipulation environment to the WIMP environment	4	5	3	3	5	3	3.83
11. It was simple to return from the WIMP environment to the VR-Manipulation environment	4	4	4	4	5	4	4.17
12. It was simple to transition from the VR-Navigation environment to the WIMP environment	5	4	4	2	5	3	3.83
13. It was simple to return from the WIMP environment to the VR-Navigation environment	4	4	4	3	5	4	4.00

Participants P2 and P6 confirmed at the interview that they preferred to use the LCD at the left side, not moving it to the center of the desk. The left-handed participant also confirmed problems to move the LCD from left to center.

Table 8 presents the answers to questions 14 to 28 of the usability questionnaire. Comfort in the use of the wand presents a satisfaction index of 3.17, or 54%, indicating that participants felt their arms tired in the prolonged use of the wand. The system designer understood that the keyboard support could also be an arm rest, but participants did not realize this, keeping their arms up at all times during interaction, causing the tiredness.

On the other hand, the satisfaction index for overall system usage was 4 in a 1 to 5 scale, or 75%, implying that the participants enjoyed using the system. The interviews confirmed this overall satisfaction.

Table 8. Answers to the Usability Questionnaire – Factors.

Usability Questions - Factors	P2	P3	P4	P5	P6	P7	Avg.
14. It was easy to memorize the commands to transition between environments	4	3	4	3	5	2	3.50
15. Animations have helped the perception of transitions	5	5	4	4	5	4	4.50
16. It was easy to keep oriented during the environment transitions	5	5	4	3	4	4	4.17
17. I did not notice interaction inconsistencies between the Navigation and Manipulation environments	5	4	2	1	4	5	3.50
18. I felt comfortable using glasses in the WIMP environment	4	4	1	3	3	1	2.67
19. I felt comfortable using the wand	4	3	3	4	3	2	3.17
20. I felt comfortable using the keyboard	5	4	4	5	5	4	4.50
21. I felt comfortable manipulating objects at the LCD	4	4	3	2	4	4	3.50
22. I felt comfortable using glasses during the interactions in general	5	3	1	1	4	2	2.67
23. I felt comfortable moving the LCD sideways	N/A	3	3	1	4	2	2.60
24. I did not feel cyber-sickness using the system	5	5	5	5	3	5	4.67
25. I felt comfortable interacting with the system	4	3	3	3	3	3	3.17
26. It was easy to learn to use the system	4	4	5	3	5	3	4.00
27. When I made a mistake, I recovered easily and quickly	2	5	4	1	4	2	3.00
28. In general, I am happy with this system	4	4	4	4	4	4	4.00

6. DISCUSSIONS

6.1 Unique Contribution of Each Method

A total of 82 HCI issues were identified by the composition of evaluation methods proposed in this study. Table 9 presents the ten issues with the highest number of occurrences.

Table 9. Summary of the Main HCI Issues.

N°	Main HCI Issues	Heur. Eval.	User Obs.	Quest. /Inter.	CEM
1	Navigation is slow when approaching the 3D scene from a distant point.	1	4	3	2
2	No *feedback* at the LCD screen when releasing the 3D icon, to indicate if it was attached to the object or not.	1	1	3	4
3	User did the inspection of the tip of the crane by rotating the whole 3D scene at the navigation environment.		4	1	3
4	Unconfortable to move the LCD monitor to the central position, preferring to leave it always by the side.		2	5	
5	User pressed the A buttom twice and returned from manipulation to WIMP.		1	3	3
6	Prolongued use of the wand causes fatigue of the users' arm.	2	1	3	
7	Users found the glasses unconfortable, especially those that already use glasses.	1	1	4	
8	Verified the tip of the crane but chose not to leave a message because there was no crack.		3	2	1
9	It was not clear how to return from manipualtion to navigation because the use of the Home buttom was not very consistent.	2		2	1
10	The 3D icon that represents the annotation is small, hard to see and to select.	1		3	1

Table 10 presents a summary of all the HCI issues identified by the usability and communicability methods.

Table 10. Evaluation Results Summary.

HCI Issues	Heur. Eval.	User Obser.	Quest./ Interv.	CEM
Total Occurrences	26	54	55	43
Distinct Issues	25	37	34	31
Unique Issues of each method	13	13	14	8

In spite of using a smaller number of participants, the communicability evaluation method was capable of identifying 8 issues not detected by the usability methods, because they did not cause a perceived usability issue, although in some cases they were on the verge of doing so. This was the case of communication expressions that only lasted for a few seconds, but provided evidence that the participant was temporarily disoriented, either having some difficulty to recognize an object, and/or not knowing what to do to proceed. These cases did not last for long and the task execution was rapidly resumed in the correct fashion, with no perceived impact on the task performance. Regarding the usability evaluation methods, each method has revealed a relevant number of unique issues. The execution of the CEM using the same user observation session, and the application of the questionnaire and interview to the same participants, has allowed a significant increase in the discovery of new issues, without much impact on the resources employed to perform a regular usability evaluation.

The knowledge about interaction design and evaluation of VE applications is still recent and lacks well established methods and techniques. This study compared the results of various evaluation methods, demonstrating that they all contribute in a distinct way. User observation with talk aloud combined with questionaries and interviews are commonly used for this kind of evaluation. They raise usability issues based, respectively, on the user observation by specialists and on the users' perception of the system. Heuristic evaluation reveals usability issues based on experts' view of the system. The most innovative and recent method in this study is CEM, which indicates communicability issues. Communicability is defined as "the property of software that efficiently and effectively conveys to the users its underlying design intent and interactive principles" [4]. Therefore, this composition of methods adds a new dimension to VE interaction evaluation: the quality of the metacommunication from the system designer to the user, as will be discussed in the following section.

6.2 User Interaction with Multiple Designers

Several of the issues were related to the exposure of the user, at interaction time, to very different signification systems, represented by very different interface components: the 3D scene (content designer), the 3D visualization application (application designer), the textual task instruction (task manager/designer), the 3D visualization hardware (MiniCave designer), the input device (Wand designer) and the operating system (OS designer). This multitude of at least six designers has been exacerbated by the interaction with Hybrid VEs, where each designer produces its own and quite different interface component. When evaluating standard WIMP applications, this multitude of designers is not so evident or pronounced.

The evaluation results provide strong evidence that the compatibility and consistency of these interface components is paramount, otherwise the user will perceive a confusing interaction experience, leading to errors, lower productivity and frustration. The application of the CEM was keen to identify the inconsistency of the signification systems produced by these multiple designers.

One of the challenges for the future will be to make these signification systems more compatible and consistent, or at least find an easier or simple way to map from one signification system to the other, like using mini-maps or a representation of the compass rose, superimposed to the 3D scene, when navigating.

There is some evidence from the results that semiotic engineering should be capable of helping to find solution to this challenge, and lead to better ways to communicate the designer's intent to the user.

6.3 Issues Related to the Input Device

Several of the issues were caused by the difficulties of manipulating the *Wand* and its buttons. This was not unexpected, because most users are not familiarized with this kind or manipulation, and much less with the specific implementation chosen by the HybridDesk designer. The WIMP environment utilizes for input the keyboard and mouse, which are considered such a standard that no special attention tends to be paid to them in HCI evaluations. But VEs demand different and in some cases unique input devices, with non-standard forms and behaviors, so a lot of attention has to focus on designing the interaction of VEs with the input device. Evaluation results have highlighted this fact and, in particular, that the CEM can provide unique results in evaluating the interaction with VE input devices.

7. CONCLUSION

This study has demonstrated that a composition of evaluation methods is important to reveal user interface and interaction problems and issues that no method alone can uncover. Moreover, it can find issues of different natures without adding much to the evaluation duration and/or cost. The addition of CEM, based on semiotic engineering, was important to find several types of issues, particularly those that result in the difference between the designer's intention and the user's perception of the system. This is important for the evaluation of complex systems, including VEs and multimodal interaction.

Due to the great complexity and lack of established knowledge on best practices of interaction techniques for VEs, the application of a composition of methods provides a much more comprehensive evaluation tool, due to the uniqueness of the findings of each method. Moreover, the proposed composition of methods was conceived in a way to minimize the impact on the overall cost of the evaluation, by combining steps of different methods. This constitutes a viable approach to provide a richer evaluation of systems with complex interactions.

This study has also highlighted the need to pay attention to the interaction of hybrid VEs with the input devices, due to its uniqueness compared to the standard interactions with keyboard and mouse.

Finally, this study has exposed the importance of the compatibility and consistency of all the signification systems to which the user is exposed at interaction time. Six different types of signification systems and designers have been identified, but there may be more. This suggests that a tightly coordinated system design is required to provide a consistent and satisfactory system behavior, especially for hybrid VE systems, otherwise the user experience will be disrupted and the product will fail.

8. ACKNOWLEDGMENTS

Tecgraf is a laboratory mainly supported by Petrobras. Thanks also to the Brazilian National Research Council (CNPq) and FAPERJ for the individual support for the authors. Alberto Raposo thanks FAPERJ (#E-26/102.273/2009). Simone Barbosa thanks CNPq (#313031/2009-6).

9. REFERENCES

[1] Bowman, D., Gabbard, J. and Hix, D. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments, 11*, 4 (2002), 404-424.

[2] Bowman, D., Kruijff, E., LaViola Jr., J. and Poupyrev, I. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2005.

[3] Carvalho, F., Raposo, A. and Gattass, M. Designing a Hybrid User Interface: a case study on an oil and gas application. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry (VRCAI '09*) (Yokohama Japan, December 2009), ACM Press, 2009, 191-196.

[4] de Souza, C. *The Semiotic Engineering of Human-Computer Interaction*. The MIT Press, Cambridge, MA, 2005.

[5] de Souza, C. and Leitão, C. F. *Semiotic Engineering Methods for Scientific Research in HCI*. Morgan & Claypool, 2009.

[6] Ericsson, K. and Simon, H. *Protocol Analysis: Verbal Reports as Data*. Revised Edition. The MIT Press, 1993.

[7] Gabbard, J. L. *A Taxonomy of Usability Characteristics in Virtual Environments*. Master Thesis, Faculty of Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, 1997.

[8] Gabbard, J.; Hix, D. and Swan II, E. User-Centered Design and Evaluation of Virtual Environments. In *Proceedings of IEEE Virtual Reality '99*, IEEE Computer Society Press, 1999, 51-59

[9] Kuniavsky, M. *Observing the User Experience: A Practitioner's Guide to User Research*. Morgan Kaufmann Publishers Inc. San Francisco, CA, 2003.

[10] Mulder, J. D. and van Liere, R. The Personal Space Station: Bringing Interaction Within Reach. In *Proceedings of Virtual Reality International Conference (VRIC 2002)*, 2002, 73-81.

[11] Nielsen, J. *Usability Engineering*. Academic Press, 1993.

[12] Octavia, J., Raymaekers, C. and Coninx, K. Adaptation in virtual environments: conceptual framework and user models. *Multimedia Tools and Applications*, 2010 (online first), doi: 10.1007/s11042-010-0525-z.

[13] Peixoto, D. C. C., Prates, R. O. and Resende, R. F. Semiotic inspection method in the context of educational simulation games. In *Proceedings of the ACM Symposium on Applied Computing (SAC 2010)*, 2010, 1207-1212.

[14] Prates, R. O.; de Souza, C. S. and Barbosa, S. D. J. A Method for Evaluating the Communicability of User Interfaces. *ACM Interactions, 7*, 1(2000), 31-38.

[15] Sharp, H; Rogers, Y. and Preece, J. *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons Ltd., UK, 2007.

[16] Tullis, T. and Albert, B. *Measuring The User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann 2008.