

Survey on Behavioral Observation Methods in Virtual Environments

Samson Habte Tesfazgi
ID No. 1031023

Supervisors

*dr. ir. C.A.P.G. van der Mast,
Delft University of Technology,
Faculty of Information Technology and Systems*

*Ing. A.O.Brugman,
Green Dino Virtual Realities
Wageningen*



**Delft University of Technology
Faculty ITS
Section Human-Computer Interaction**



**Green Dino Virtual Realities
Wageningen,
The Netherlands**

April 2003

Contents Table

Contents Table	2
Chapter One: Introduction	3
1.1 Research Problem Description.....	3
1.2 Research Approach.....	5
1.3 Report Structure	5
Chapter Two: VR Technology	6
2.1 What is Virtual Reality	6
2.1.1 Virtual Environment	7
2.1.2 Virtual Reality Interfaces.....	9
2.1.3 Modes of Interaction in VR System.....	10
2.2 Virtual Reality Applications	13
2.3 Current Challenges.....	14
2.4 Concluding Remark	16
Chapter Three: Behavioral Observation Methods in Real Environment	19
3.1 Introduction.....	19
3.2 Behavioral Observation	19
3.2.1 Self Report	21
3.2.2 Direct Behavioral Observation	23
3.2.3 PsychoPhysiological Recordings.....	25
3.3 Computer Based Behavioral Observation.....	25
3.3.1 The Observer.....	26
3.3.2 CAMERA	30
Chapter Four: User Behavior Observation Methods in Interactive Computer Applications	31
4.1 Introduction.....	31
4.2 Usability Evaluation Methods.....	31
4.2.1 Direct Observation	32
4.2.2 Interview and Focus Group.....	33
4.2.3 Questionnaire	33
4.2.4 Performance Measuring	35
4.2.5 Logging User Actions	35
Chapter Five: Discussion on Insights Gained from Related Domains.....	38
5.1 Insights from Behavioral Observation Methods in Real Environments	38
5.2 Insights from User Behavior Observation Methods in Interactive Computer Applications	42
Chapter Six: Current Developments Toward Observation Methods in VEs.....	46
6.1 Human Performance Measurement in Virtual Environments.....	46
6.2 Measuring Presence in Virtual Environments	47
6.3 Usability Evaluation Techniques	48
Chapter Seven: Conclusion and Recommendations.....	50
7.1 Conclusion	50
7.2 Recommendations.....	52
References.....	53

Chapter One: Introduction

As part of the requirements for the completion of the Master of Science in Computer Science at Delft University of Technology, an individual research assignment has to be conducted. This report paper is dedicated for the individual research assignment conducted at Green Dino Virtual Realities, a widely oriented engineering-office specialized in 3D visualization and simulation on the basis of VR technology since 1992.

Currently, a project named Virtual Observer is initiated at Green Dino Virtual Realities. The main goal of this project is to design and develop a software tool for observing and assessing behaviors of end users or other autonomous entities in Virtual Environments. The design and development of a behavioral observation method in Virtual Environments is to be conducted in my Master thesis project. However, a literature review research study, which surveys methods of behavioral observation in Virtual Environments, is conducted before the commencement of the Master thesis project. The purpose of the literature review study is to lay down a foundation for the Master thesis project.

1.1 Research Problem Description

Computing Science has been playing a significant role in realizing what seems to be only a dream of science fiction writers. Science fiction writers have already been exploring the role of computers in future world, and have described a synthetic 3D universe that is as believable as the real physical universe [Vince 95]. To realize such visions, users should be able to interact and explore computer systems using multi-modal interaction techniques similar to the way human beings explore real objects. To this end, a new technology has been emerging because of the underlying development of real time computer graphics and simulation software. This new promising technology is Virtual Reality (VR). VR systems exploit a mode of interaction that becomes possible with immersive real time computer graphics systems.

Virtual 3D objects can be viewed from any direction; integrated with other objects in to an assembly; the surface properties can be changed, it can be weighed; and it can even be stressed until it breaks with the help of Virtual Reality technology by taking users a part of the virtual domain [Vince 95]. This is the ideal goal of Virtual Reality technology. In today's virtual environments, users are immersed into an experience with sub-optimal visual resolution, inadequate spatialization of sound, encumbering interactive devices, and misregistration of tracking information [Kay & Stanney 2002]. There fore, a lot of issues regarding technological and scientific challenges have to be resolved in order to realize the full potential of VE technology. One aspect of these issues is concerned in how best to design VE applications and use them effectively and productively, for instance issues concerning methods useful for conducting usability evaluation (formative usability evaluation in which representative users are observed while performing a prescribed scenarios of tasks) and other relevant behavioral analysis in virtual environments.

Virtual Reality technology has been applied in various domains of application. Some of these domains of application are engineering, medicine and health care, training, education, entertainment, television, scientific visualization, and military training. The most popular and promising applications are training simulators, ergonomics assessment of prototypes, and reduction of phobic reactions like fear of height and flying. In all of these applications, end users interact via multi-modal interaction in an immersive or semi-immersive environments. During a specific usage of VR applications, end users issue commands; navigate the virtual world; and manipulate the virtual world. In other words, end users respond to various stimuli generated by a virtual world and the virtual world senses various stimuli like head and hand movements and reacts in an appropriate way.

Recording of how end users or other autonomous entities behave in virtual environment during a certain session can have valuable advantages, for instance, in evaluating the usability of VR application prototypes and assessment of behavior of end users or other autonomous entities in a certain Virtual Environment. The behavior of end users or other autonomous entities in Virtual Environment can be observed and assessed for various reasons, for instance to measure how trainees perform in a VE-based training simulator. Trainees need to get feedback of how they perform and want to know how they progress. Measuring the performance of end users of VE-based training simulators is also important for assessing skills of trainee for personnel selection purposes.

Behavioral researchers can conduct observational studies with the help of VR technology in situations that are too expensive or time-consuming to be reproduced in the real world, or in situations that are potentially harmful for the persons involved. It is obvious that behavioral researchers need to record and evaluate how participants behave in the specific Virtual Environment created to simulate an existing real environment or non-existing environment.

VR technology can be used also for virtual prototyping of new products and services in which the interactive and immersive environment is effective for ergonomic assessment of the newly designed product or service, for identifying design flaws, and generating new ideas at an early stage. To assess the ergonomics of the prototypes, recording how the potential user of the designed product or service behaves when he/she navigates and interacts with the prototype is necessary for identifying potential ergonomic problems.

The need for recording how end users or autonomous entities behave in Virtual Environment during a specific usage or session for reasons mentioned in the above, poses the following main research question that is to be addressed in this research study:

- *Which technique or method is the best for conducting behavioral observation in Virtual Environments?*

1.2 Research Approach

Virtual Environment as its name implies is a virtual representation of an existing or non existing physical environment or an abstract information which offers end users real time interactivity and make them feel as if they are part of it. Due to the interactive and immersive nature of Virtual Environments, behaviors that occur in real environments have resemblance with behaviors that occur in Virtual Environments. On the other hand, Virtual Reality applications have similarity with computer systems that are interactive such as multimedia and web applications even though the degree or scope of interaction is more in Virtual Reality applications.

Based on the above facts, the main research question is approached by discussing four major issues. The first issue is concerned with understanding and defining how users interact in Virtual Environments, exploring challenges of Virtual Reality technology, and justifying the societal and scientific relevance of the outlined research problem. The second issue is concerned with what and how techniques and concepts regarding observation and assessment of behaviors can be applied to Virtual Environments from other relevant areas: real environment and interactive computer systems. The third issue is concerned with assessing the current developments in techniques to record and assess the behavior of end users or other autonomous entities in Virtual Environments. The fourth issue is regarding what technique or method is the best to record and assess behaviors of end users or other autonomous entities in Virtual Environments.

1.3 Report Structure

The structure of the remaining part of this report paper is as follows. In chapter two, the overview of Virtual Reality technology is presented. In chapter three, observation methods of behavior in real environment is described in general. In chapter four, observation methods of user behavior in interactive computer systems are discussed. In chapter five, the application of insights gained from behavioral observation methods in real environment and interactive web applications into Virtual Environment is discussed. In chapter six, current developments toward observation methods in Virtual Environments are discussed. Finally, conclusion and recommendations of the best technique for observing and assessing behaviors of end users or other autonomous entities in Virtual Environments is given.

Chapter Two: VR Technology

In this chapter, the general overview of Virtual Reality technology and its applications are described. Its purpose is to have insight into the way users interact or behave in Virtual Environments and to justify the scientific and societal significance of the research problem to be studied in this research paper.

2.1 What is Virtual Reality

The large number of publications, TV shows and conferences that suddenly appeared, described Virtual Reality in various and sometimes inconsistent ways [Grigore & Philippe 94]. The reason for having more than one definition of Virtual Reality is due to the fact that it is a new technology and writers attempt to define it based on different perspectives such as the tools it uses, its functions, etc.

If we consider the perspective based on functionality, Virtual Reality is a simulation in which computer graphics is used to create a realistic looking world. Moreover, the synthesized world is not static but responds to user inputs (gestures, verbal commands, etc.). This defines a key feature of Virtual Reality, which is real-time interactivity. Here real time means that the computer is able to detect user inputs and modify the virtual world instantaneously. Interactivity and its captivating power contribute to the feeling of immersion, of being the part of the action on the screen that the user experiences. The sense of immersion is even further reinforced using all human sensorial channels. A definition that summarizes the above ideas is: [Grigore & Philippe 94]

Virtual Reality is a high-end user interface that involves real time simulation and interaction through multiple sensorial channels. These sensorial modalities are visual, auditory, tactile, smell, taste, etc.

Another definition given in the book entitled “Handbook of Virtual Environments” is as follows: [Kay & Stanney 2002]

Virtual Environment (VE) is a three-dimensional data set describing an environment based on real-world or abstract objects and data. Usually virtual environment and virtual reality are used synonymously, but some authors reserve VE for artificial environment that the user interacts with.

Now that Virtual Reality is defined considering different perspective, one might wonder how VR technology is introduced to the world. Virtual Reality is not a new invention, as most people think, but dates back to more than 30 years. In 1962, Morton Heilig invented the first virtual reality video arcade entitled “Sensorama Simulator”. This early virtual reality workstation had 3-D video (obtained with a pair of side-by-side 35 mm cameras), motion, color, stereo sound, aromas, wind (using small fans placed near the user’s head) and a seat that vibrated. [Grigore & Philippe 94]

Heilig has managed to create passive virtual experiences, which were to be made interactive with the help of computers, particularly the development of computer graphics [Vince 95]. After the invention of Sensorama Simulator, several research studies have been conducted till now in an attempt to realize the ideal goal of Virtual Reality technology, which is creating a virtual 3D object in which users can interact in the same way human beings explore or interact with real things and users believe they are part of the virtual environment.

We can categorize Virtual Reality systems into three groups depending on the degree of immersion and interactivity. These three groups are immersive systems, non-immersive systems, and hybrid VR systems. Immersive systems replace our view of the real world with the computer-generated images that interact to the position and orientation of the user's head. A non-immersive system on the other hand, leaves the user visually aware of the real world but able to observe the virtual world through some display device such as graphics workstation. A hybrid VR system permits the user to view the real world with virtual images superimposed over this view – such systems are also known as 'augmented reality' systems. A practical example is found in the HMDs used by fighter pilots, which allow the pilot to view their outside world simultaneously with overlaid synthetic graphics. [Vince 95]

A generic immersive VR system consists of four system elements interacting with each other to make the whole functioning system. These four elements are the virtual environment, the computer environment, VR Interfaces, and modes of interaction. The VE covers ideas such as model building, introducing dynamic features, physical constraints, illumination and collision detection. The computer environment includes the processor configuration, the I/O channels, the VE database and the real-time operating system. VR Interfaces encompass the hardware used for head tracking, image display, sound, haptics, and hand tracking. Modes of interaction involve hand gestures, 3D interfaces and multi-participant systems. In the following three sections, the virtual environment, VR Interfaces, and modes of interaction are described in detail. The connectivity between the four system elements is depicted in figure 1.1. [Vince 95]

2.1.1 Virtual Environment

The VE can take many forms; for example it could be realistic representation of some physical environment such as the interior of a building, a kitchen, or even an object such as a car. It could be that the VE does not have any physical basis at all. For instance, it might be a 3D database of a geographical, hierarchical network describing a multinational company or a multidimensional data set associated with stock transactions. Whatever the nature of the underlying data, a geometric model is required to represent atomic entities and their relationships with one another. Based on this geometric model, a geometric database must be built to represent the environment, and stored such that it can be retrieved and rendered in real time when required. The database storing VE includes 3D geometry, color and texture, dynamic characteristics, physical constraints, and acoustic attributes. In general objects within a VE can be divided into two groups: static and

dynamic. Some dynamic objects may be defined without any constraints being placed upon their special behavior, whereas others may be physically constrained to move within prescribed limits. [Vince 95]

The basic user input signals consist of the position and orientation of the user's head and hand. The user's head position and orientation are vital input signals to the VR system, as these determine the viewpoint from which the virtual world is rendered whereas the user's hand position and orientation are used to interact with the VE. For instance, when an interactive glove is used, hand gestures can be used to initiate or terminate system processes. Similarly, a glove or a 3D mouse can be used to pick virtual objects and move them to different positions. [Vince 95]

The output signals from the VR system carry visual, audio, and haptic information about the VE, and are supported by three independent subsystems working in parallel. The primary output from the system is visualization of the VE displayed upon the user's HMD. The audio signal is derived from the acoustic subsystem for simulating sound events in the VE. These signals can be preprocessed to incorporate information that allows the user to localize the sounds in space using headphones. The haptic channel carries signals describing forces to be transmitted back to the user's body. [Vince 95]

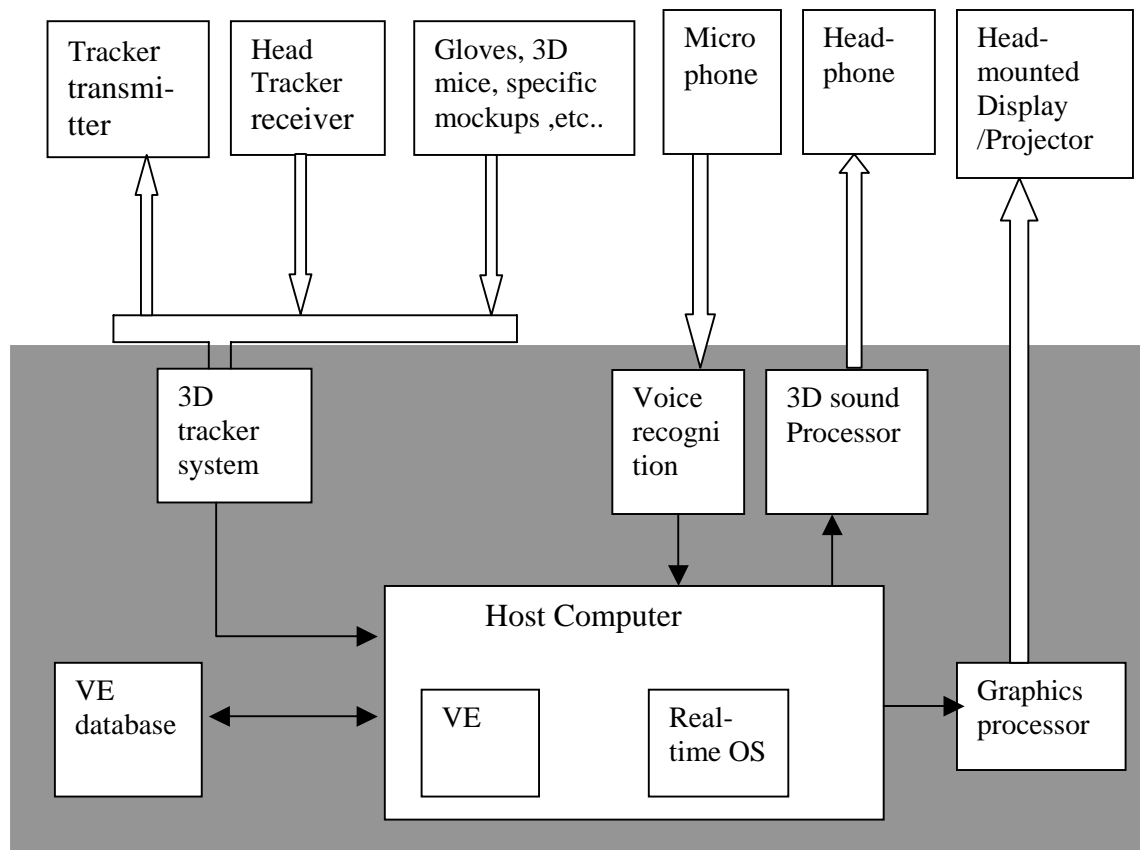


Figure 1.1 The integration of the various elements of a generic VR system

2.1.2 Virtual Reality Interfaces

One of the key features of Virtual Reality is interactivity. In order to allow human-computer interaction it is necessary to use special tools designed both to allow input to the computer and to provide feedback to the user. Today's VR interfaces are varied in functionality and purpose as they address several human sensorial channels. Here only those VR interfaces, which are popular and most used, are described. [Grigore & Philippe 94]

To generate the VE's images according to the movements of the user's body, it is fundamental to rapidly acquire data about the different body positions in the 3D space, and transmit them to the computer in order to elaborate the environment's modifications to be issued in response to the user's actions. This can be done by dedicated devices named 3D trackers. The technologies used predominantly in 3D tracking are four: mechanically based, optically based, magnetically based, and acoustically based. Mechanical tracking makes use of a mechanical armature with one side connected to the top of a helmet and the other end connected to an encoding device on the ceiling. As the user changes head position, the helmet moves the upper device like a mechanical arm, and data related on that movement is relayed to the computer. While mechanical position/orientation tracking is the most precise method of tracking, it has the disadvantage of being the most limiting. [4]

Optical tracking makes use of small markers on the body, either flashing infrared LED's or small infrared-reflecting dots. A series of two or more cameras surround the subject and pick out the markers in their visual field. Software correlates the marker positions in the multiple viewpoints and uses the different lens perspectives to calculate a 3D coordinate for each marker. One limitation of these systems is the processing time needed to analyze the several camera images and determine each marker's 3D position. Magnetic tracking uses a source element radiating a magnetic field and a small sensor that reports its position and orientation with respect to the source. Magnetic systems do not rely on line-of-sight observation, as do optical and acoustic systems, but metallic objects in the environment will distort the magnetic field, giving erroneous readings. However, the current technology used for magnetic tracking is quite robust and widely used for single or double hand tracking. Acoustic trackers use high-frequency sound to triangulate a source within the work area. Most systems like those from Logitech and the one used in the Mattel Pow pings from the source received by microphones in the environment. [4]

Handheld 3D mice are used to control VR object's position. The position and orientation of the 3D mice is monitored similarly to the method used for the user's head. Several push buttons are provided on the 3D mice, within reach of the user's fingers. These buttons are binary on/off and can be programmed according to the application. In order to have large gesture based interaction with the VR simulation, it is necessary that the I/O tools maintain the hand freedom of motion within a certain volume. It is also desirable to allow additional degrees of freedom by sensing individual finger motions. Sensing gloves are VR tools that are designed to address the above requirements. [Grigore & Philippe 94]

The typical sensing glove used in Virtual Environments is the “data glove”, which is somehow like a hi-tech hand-shaped pointing device. It can be used to initiate commands and to navigate through the VE. One problem with the data glove is the need to provide a sense of touch to increase the haptic experience of the wearer. Commercial devices to add touch to virtual worlds have only the most rudimentary capabilities at present. These devices can be broadly separated in tactile feedback, force feedback, thermo feedback. [4]

Human vision is the most powerful sensorial channel and has an extremely large processing bandwidth. It is therefore not surprising that there is a multitude of visual feedback tools for VR, the most common being the head-mounted displays (HMD). Other visual feedback systems are “booms,” stereo, “active” glasses and monitors and large-screen stereo projectors. [Grigore & Philippe 94]

An HMD isolates the user from the real world and substitutes binocular views of the VE. One approach to HMD design involves two small flat-panel liquid crystal displays, in front of which are mounted optics to provide a comfortable focal point. In the real world our eyes are constantly converging and focusing as our attention is drawn from one point of interest to another. This is not possible with this type of HMD, which is fixed-focus and the natural eye activity we have mastered in the real world cannot be done. Further technology is available to track the orientation of the user’s eyes. [Vince 95]

Highly immersive virtual reality simulations should have “virtual audio” in addition to stereo graphics [Grigore & Philippe 94]. Virtual sound is a recorded sound experience that contains significant psycho acoustic information to alter human perception into believing that the recorded sound experience is actually occurring in reality [Grigore & Philippe 94]. One possible technical way to provide virtual audio is via transducers positioned at the entrances to the ear canals (headphones). Headphones are fixed to the head and thus move simultaneously with it. [4]

2.1.3 Modes of Interaction in VR System

VR systems are very different from other computer systems because of their immersive and interactive feature. Information can be exchanged between end-users and VR systems via several human sensorial channels. These facts create an opportunity to have various interaction modes or paradigms that can be realized with the VR interfaces discussed in the previous section.

Before we discuss the different modes of interaction in VR system, the general model of the interaction between human users and the VE is described based in previous research studies. The human output channels and computer input modalities define the input flow into the Virtual Environment. The computer output modalities and human input channels define the feedback flow by which the user receives feedback from the virtual environment. These two communication channels form the human-VE interaction loop. The processes involved at the human side of the interaction loop are perception,

cognition, and control. Perception is the process of transforming sensorial information to higher level representations. In the model adopted here, perception involves the machine-to-human communication flow. The corresponding perception channels are visual, auditory, and somatic (haptic, olfactory, gustatory, vestibular). The control process, in the context of VE systems, represents the translation of human actions into task-related information for the computer. Thus control involves the human output channels and computer input modalities. The control modalities of gesture, speech, gaze, and touch are used in conjunction with hardware interfaces. The Human-VE interaction model is represented schematically in the following figure. [Kay & Stanney 2002]

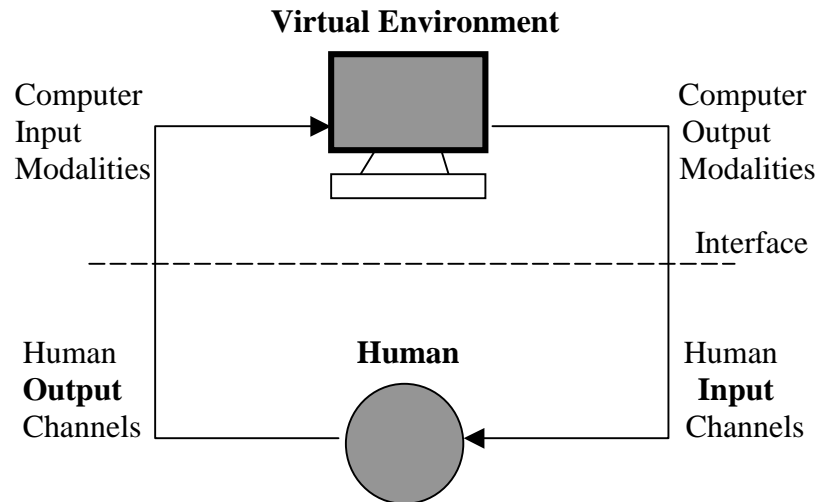


Figure 1.2 Human-VE interaction model adapted from Schomaker et al., 1995

Development of multimodal input for VE systems is essential due to the increased user control and interactivity required by recent VE applications. Human input in VE is mostly achieved through gesture-controlled input devices. The most prevalent input modality used in present VE system is movement of the user's body. The efficient use of gestures resides in their integration with speech input. Another approach to help speech is to add gaze input. [Kay & Stanney 2002]

Interaction modes are implemented, in one way or another, in almost any modern machine or piece of equipment. They represent the different behaviors, or functions of a given system. User behavior defines what the user perceives the environment and how the interaction with the VE is progressing. Although there are a variety of different interaction techniques, many can be grouped loosely into the interaction mode that they apply. Some examples are: [Shamus & David 1]

- Navigation: walking, flying, worlds in miniature, head-butt zoom, two handed flying
- Selection: laser pointing, cone-casting, head crusher select, go-go interaction, look-at menus
- Manipulation: scaled world grab, orbital mode, hand-held widgets, over-the-shoulder deletion

- Environment commands: pull-down menus, voice recognition, gesture recognition.

In the above, several interaction techniques applicable in VR systems are mentioned and are listed under the mode of interaction they apply. Here, how interaction techniques are defined generally and the description of the interaction modes mentioned in the above are presented. There are three general areas that are of interest when examining interaction techniques, namely, their specification, modeling, and evaluation [Shamus & David 2].

Specification of interaction technique includes defining what the technique provides and what the technique requires. Interaction techniques provide elements both to the user and system and interaction technique requirement can be system, technology or user based. Techniques for modeling interaction to address specific design questions is an active area of research, even without considering the apparent complexity of VR. The very nature of virtual environments makes them difficult to describe and model. Typically, they are extensively visually based, have a non-linear process and control flow. [Shamus&David 2]

The most common task of all in VEs is that of navigating through the space of the environment. Perhaps the most natural method for navigation, though not the simplest to implement, is to use physical user motion that can be implemented with treadmills, roller skates, bicycles, etc. This method contains inherent constraints, in that the user can move freely only in two dimensions. Another simple technique that is often implemented is artificial flying, usually in the direction of the user's gaze or the user's pointing. Generally, the user simply looks or points in the direction he/she wants to go, and presses a button or makes a gesture to fly in that direction with constant velocity. One problem in using flying technique is that the user can easily get lost or disoriented if given complete freedom. To solve this problem, walking can be introduced. This is the same as the flying technique, except that the user's head is constrained to a given height above the ground plane. [Doug & Larry]

A more complex technique for navigation purpose is the head-butt zoom. It uses both hand and head motion to control the interaction. Initially, the user forms a view rectangle using their hands to mark the positions of the top-right and bottom-left corners of a rectangle. Once this is completed, head motion is used to switch between environment views.[Shamus & David 2]

Another universal task in most virtual environment applications is the selection of objects in the virtual world. The most obvious and trivial technique for object selection is simply to select an object when the user's hand comes into contact with it. However, this technique has drawbacks. The user must physically navigate to distant objects in order to select them and the selection device (the hand) is not precise enough for the differentiation of small densely crowded objects. Extending the desktop "point-and-click" idea to a VE, several applications use ray-casting for object selection. In this scheme, a ray is extended from the user's hand into the environment, and the first object intersected

by the ray is selected. Cone-casting, in which objects within a narrow cone cast from the user's hand is selected, is also used for selection task. [Doug & Larry]

A different technique for selecting objects in virtual world is the head crusher select, which is based in an immersive virtual environment and involves operating on 3D objects at a distance by interacting with 2D projections on the image plane. [Shamus & David 2]

Manipulation is the task performed after objects are selected. It might include movement transformation, modification of selected objects, and creation of new objects. It also include those actions which are more symbolic or abstract in nature such as loading of files and changing system parameters. There are basically two choices for most object manipulation tasks. These are direct manipulation and indirect (command-driven manipulation). Another universal task in virtual environments is the issuance of commands by the user. It is proved that not all tasks are suited for direct manipulation or for gestures. If a completely abstract, symbolic task is accomplished in one of these ways, the manipulation of gesture, in the end, is simply another command with its own syntax. One popular method of issuing commands in a VE is through voice-recognition. Menu-based commands can be also used for issuing commands in VE. It has been also implemented a hybrid technique that uses voice-recognition and menu-based commands. [Doug & Larry]

2.2 Virtual Reality Applications

The virtual domain of computers offers reliability, speed, ease of access, compactness and security, and is easily transmitted to other computers located in distant parts of the world [Vince 95]. It is easy and cost-effective to manipulate a virtual domain than its counterpart real domain. Due to these facts, VR technology has been a promising technology applicable in various domains of application. Here the most popular applications of VR technology are described to have overview on its wide application areas. These most popular domain of application are training simulators, medical and health care, education, defense, engineering, ergonomics and human factors research, database and scientific visualization, and entertainment industry. The most widely use of VE technology is in the entertainment market due to the emergence of VE video games both in amusement arcades and at home on video games consoles [Kay & Stanney 2002].

Training simulators are used for planes, submarines, power plants, tanks, helicopter, ships, cranes, trains, surgery, endoscopes and air traffic control. Such simulation uses a replica of the real operational environment and real time computer to model its dynamics. Training through simulation provides significant benefits over other methods. For instance, a hazardous environment, such as a nuclear power station, or an aircraft landing in a fog can be accurately simulated without any danger to the trainee. Other benefit is the ability of computer software in providing flexibility to structure training programs and even monitor and measure the progress of a training session. Many simulators employ

computer generated images. However, the emergence of immersive VR system is providing new training paradigms for training applications. [Vince 95]

The key components of medical and health care where Virtual Environment can be applied are diagnosis, therapy, education and training, and medical records. Diagnosis using virtual endoscopy is one of the areas that can achieve clinical efficiency in the earliest time frame. Virtual Environments can be used in computer assisted surgery, image guided surgery, tele- surgery, and treatment of phobias and other psychological disorders. [Vince 95]

Virtual Environments can be used for learning of the kind expected to occur in schools, colleges and universities, that is, the acquisition of general problem solving skills, mastery of facts and concepts, and improvement of the learning process itself. VE applications in the defense area are used primarily in large scale simulators designed for such activities as operations planning, war gaming, command-control-communications and intelligence, tri-service pilot, navigator, and driver training. [Kay & Stanney 2002]

The applications of VE in the engineering domain are many and varied, they cover heavy engineering sectors of petrochemical and hydroelectric plant design or construction sites in general, to the ergonomic and aesthetic evaluation of automobile and civil aircraft interiors, to the prototyping of nuclear control systems. [Kay & Stanney 2002]

VE can help scientists to visualize all manner of complex artifacts and processes, be they new protein chains, viral compositions, chemical drug concepts, or gaseous particles in motion. It can also be applied to visualize and explore the atomic surfaces of angstrom-sized material samples. As NASA, the United Kingdom's ARRC, and related establishments adopted a human-centered approach to the use of VE to control macro-sized robots for hazardous environments, future micro-robots and even nanobots – such as the DNA “screw driver”, designed for deployment within the human body-will need strict supervision by skilled operators equipped with advanced VE equipment. [Kay & Stanney 2002]

2.3 Current Challenges

Virtual Environments are driven by the technology that is used to design and build these systems. This technology consists of the human-machine interface devices that are used to present multimodal information and sense the virtual world, as well as the hardware and software used to generate the virtual environment [Kay & Stanney 2002]. Beside the technological challenges, other kind of challenges are also existing such as psychological consideration and limitation of traditional usability evaluation methods for assessing VE systems. These challenges have been encountering constraints that hinder the progress of Virtual Environments in the aspect of cost, performance, and usability. In the following paragraphs, the technological, psychological, usability evaluation challenges are discussed.

When an input device is connected, such as a body or limb tracker, a serial port is generally utilized, a port typically designed for character input and not high-speed data transfer. A solution to the input device connectivity issue that is available on commodity computing is the great unsolved problem. Visual displays, especially head-mounted displays have come down substantially in weight but are still hindered by cumbersome designs, obstructive tethers, suboptimal resolution, and insufficient field of view. Recent advances in wearable computer displays, which can incorporate miniature LCDs directly into conventional eyeglasses or helmets, should ease cumbersome design and further reduce weight. Virtual Retinal Displays (VRDs) may bring truly revolutionary advances in display technology. VRD technology, which was invented in 1991 at the University of Washington's HIT Lab, holds the promise for greatly enhanced optics. With technology, an image is directly scanned into viewer's retina using low-power red, green, and blue light sources, such as lasers or LEDs. The VRD system has superior color fidelity, brightness, resolution, and contrast compared to LCDs and CRTs, as it typically uses spectrally pure lasers as the light source. With advances in wireless and laser technologies and miniaturization of LCDs, during the next decade visual display technology should realize the substantial gains necessary to provide high-fidelity virtual imagery in a light weight noncumbersome manner. [Kay & Stanney 2002]

Even though, there has been a tremendous gains in the area of virtual auditory displays, for instance, systems are available that present multiple sound sources to multiple listeners using positional tracking, technology for designing complex spatial audio scenarios, including numerous reflections, real-time convolution, and head tracking is not available currently. Current haptic technology provides net force and torque feedback but has yet to develop effective tactile feedback. . [Kay & Stanney 2002]

There are a number of psychological considerations associated with the design and use of VE systems. Some of these focus on techniques and concerns that can be used to augment or enhance VE interaction and transfer-of-training, while others focus on adverse effects due to VE exposure. In contrast to the limited knowledge concerning perceptual and cognitive design principles that augment or enhance VE interaction, more is known about identifying and controlling the adverse effects of VE exposure. These adverse effects are both physiological and psychological. A survey of the current state of knowledge regarding physiological aftereffects concluded that the two most critical research issues in this area are standardization and use of measurement approaches for VE aftereffects, and identification and prioritization of sensorimotor discordances that drive these aftereffects. Taken together, the research into psychological considerations of VE exposure indicates that more research is needed to derive perceptual and cognitive design strategies that enhance VE interaction, and that there are risks associated with VE exposure. However, usage protocols have been developed that, if successfully adopted, can assist in minimizing these risks. [Kay & Stanney 2002]

Most VE user interfaces are fundamentally different from traditional graphical user interfaces, with unique I/O devices, perspectives, and physiological interactions [Kay & Stanney 2002]. Thus traditional usability engineering methods are not well suited to VE systems. Currently, very few principles for the design and evaluation of VE user

interfaces exist. Previous research studies have identified limitations of existing usability methods for assessing VE systems and suggest that novel aspects of human-VE interaction require new evaluation tools [Kay & Stanney 2002]. Toward this end, Gabbard and Hix (1997) have developed a taxonomy of VE usability characteristics that can serve as the foundation for identifying usability criteria that existing evaluation techniques fail to fully characterize [Kay & Stanney 2002]. Stanney, Mollaghasemi, Reeves (2000) used this taxonomy as the foundation on which to develop an automated system, MAUVE (multi criteria assessment of usability for Virtual Environments [Kay & Stanney 2002]. Similar to the manner in which traditional heuristic evaluations are conducted, MAUVE can be used at various stages in the usability engineering life cycle, from initial storyboard design to final evaluation and testing and used also to compare system design alternatives [Kay & Stanney 2002].

2.4 Concluding Remark

In this chapter, the basic definitions and concepts related to Virtual Reality technology are presented. The different components of a Virtual Reality system is described and the relationship among the component elements that make up the whole VR system is discussed and represented in a diagram. One of the main purposes of this chapter is to have insight into how users can interact with a VR systems. Here, the insight gained from the previous sections is discussed.

The user interaction can be represented in different levels. In the lowest level, the user interaction can be represented as a flow of signals from the human output channels to the VE system input modalities and a flow of signals from the VE system output modalities to the human input channels. During the interaction process, the user of VE systems perceives visual, audio, haptic, and other signals from the virtual environment and responds by doing actions like gesture, speech, gazing, and touch.

The other high level representation of the interaction of users with VR systems is based on the kind of task or activity the user can do while interacting with the VR system. Generally, any user of VE systems can do four kinds of tasks, which are navigation, selection, manipulation, and issuance of command. Each task is a sequence of the low level events such as flow of visual, audio, and haptic signals, and low level actions such as gesture, speech, gazing and touching. The description and order of the subtasks of the navigation, selection, manipulation, and issuance of command depends on the type of the selected interaction technique. Each general task (navigation, selection, manipulation, and issuance of command) can be realized with several interaction techniques as it is mentioned in the previous section. For instance, in artificial flying interaction technique used for navigation task can be described as: the user simply looks or points in the direction he/she wants to go, and presses a button or makes a gesture to fly in that direction with constant velocity.

Several application domains of VR technology are described and each domain has its own purpose. The semantic or meaning and purpose of each general task depend on the type of the VR application under consideration. For example, navigation task in driving simulator means driving in the traffic environment where as it means flying over sky and landing on an airport in the case of flight simulator.

The second purpose of this chapter is to justify the societal and scientific relevance of the research problem to be addressed in the research assignment and master thesis project. In the previous section, it has been discussed the challenges of Virtual Reality systems in the aspect of technological, psychological, and usability issues. The research problem to be addressed in this paper has relevance to the challenges related to psychological and usability issues.

One of the challenges concerning these two issues is to enhance the design and use of Virtual Reality applications. To enhance design of VR applications, researchers are attempting to develop an integrated usability evaluation method well suited for VR applications. Formative usability evaluation technique is an observational evaluation method that ensures usability of interactive systems by including users early and continually through out user interface development. In the past few years, a first hand evidence has been seen indicating that formative evaluation process is also an efficient and effective method of improving the usability of VE interfaces [Kay & Stanney 2002]. The necessary steps taken during formative evaluation are development of user task scenarios, identifying representative users, collecting both qualitative and quantitative data, and assessment of the collected data. The research issue addressed in this paper, which is the design and development of observation method of user behavior in Virtual Environment, will contribute a lot in the collection and assessment of data during formative evaluation process.

The other relevance of the design and development of observation method of user behavior in Virtual Environment is during the use of Virtual Reality applications. In some Virtual Reality applications, the user interacts with the system in a dynamic and complex manner and it is required to know how the interactions in the Virtual Environment progress. For example in training simulators, it is useful to know and record how a trainee behaves in a training session in order to monitor and assess the performance of the trainee.

Development of observation method of user behavior in Virtual Environment can enhance also the use of VR technology for ergonomic and behavioral researchers. Due to the interactive and immersive nature of VR applications, ergonomic researchers can evaluate the ergonomics of a newly designed product or service by creating an appropriate virtual environment where a potential user of the product or service experiments with it. In this regard, an observation method will have a great contribution in collecting ergonomic data during the experimentation and assessing the collected data to evaluate the ergonomics of a new product or service.

Directly assessing a client's behavior in the natural environment would seem to be the most valid approach to information gathering. However, direct observation in the natural environment can be quite costly, time consuming, and harmful to the client. In such situations, analogue assessment offers a practical alternatives. Traditionally, analogue assessment is done by providing the stimulus events in written or audio or video form. Nowadays, the introduction of Virtual Reality technology is becoming the promising way to conduct analogue assessment. VR technology offers a reliable alternative for behavioral researchers to conduct observational studies in situations that are too expensive and time consuming to be reproduced in the real world or in situations that are potentially harmful for the persons involved. In such cases, observation method plays a significant role in collecting and assessing how the participant subjects behaves in the virtual world created for a certain behavioral research purpose. One of the research purposes can be testing hypothesis statement regarding behaviors. Testing hypothesis in Virtual Environments is very easy and effective since it is easy to manipulate the environmental settings and stimulus events of the Virtual Environment. Behavioral psychiatrics are also using Virtual Environments for treatment of psychological and physiological disorders such as fear of height or fly. In such cases, behaviors can be collected and assessed to evaluate the effectiveness of treatment strategies.

Chapter Three: Behavioral Observation Methods in Real Environment

3.1 Introduction

The way human beings behave in a certain real environment has similarity in some aspects with the way users of VE systems behave while interacting with the system. In both ways, there are three main processes: perception, cognition, and initiating control actions. However, the kind of information that a user of most VE systems perceive is constrained by the capability of the VE system output modalities. Currently, a user of most VE system can perceive visual, audio, and force and torque feedback kind of information. Again, the kind of controlling action activated by the user of most VE systems is constrained by the capability of the VE system input modalities (gesture, speech, touch, and gazing). In some training simulator virtual environments like the driving simulator, the user can do any activities using the cabin like mockup what he can do in real vehicle.

Because of the fact that user behavior is similar to human behavior in real environment, it is worth of studying how human behavior is observed and assessed in real environment in order to apply knowledge gained from real environment to Virtual Environment. As a result, this chapter is dedicated for studying human behavior in real environment and applying the gained knowledge into Virtual Environment.

3.2 Behavioral Observation

Before behavioral observation is described, it is better to give an overview of behavior in general. Behavior can be defined in different ways by different kinds of researchers. Here a neutral definition is given from a dictionary. Behavior is defined as the action or reaction of something under specified circumstances. Broadly, behaviors are grouped into two: overt and covert behaviors. Overt behaviors are actions that other people can directly observe. Aspects of overt behavior are activities, movements, posture, gestures, positions, facial expressions, speech, and social interactions. Covert behaviors are things we do that cannot be directly observed such as cognitive processes (remembering, thinking), feelings (emotions), and physiological responses (blood pressure, temperature).

Behavior can be also categorized into three depending on what sensorial channels the actions and reactions happened. These three channels of behavioral activity are overt-motor behavior, psychological-emotional responses, and verbal-cognitive behavior. [Anthony & Karen 86]

Behavioral observation is the collection of behavioral data while subjects engage in various activities under a certain situation and the assessment of the collected behavioral data [5]. The collected behavioral data can be assessed for different purposes or goals: describing behavior, designing or selecting behavioral correcting strategy, and evaluating

the selected behavioral correcting strategy. In the following, a detailed explanation of each purpose is given.

1. **Describing Behavior:** behavioral observation can be conducted to describe behavior, that is, to know how a certain person actually behaves under a specific circumstances. This kind of assessment is done usually to identify specific behaviors, as well as the variables (both antecedents and consequences) controlling these behaviors, that need attention or correcting measures. A comprehensive guides exist for describing behavior and its key controlling factors. One of the approaches for describing behavior is that of Goldfried and Sprafkin (1974), which is the S-O-R-C model. The S (stimulus) refers to the antecedent events (event coming before a behavior) or discriminative stimuli. The O (organism) refers the biological condition of the person as well as individual differences due to past learning experience. The R (response) refers to the various behaviors in question. The C (consequences) refers to the punishing or reinforcing events that follow the behavior. [Anthony & Karen 86]
2. **Design or Selection of Behavioral Correcting Strategy:** this kind of assessment is called functional behavioral assessment. It is a systematic process for gathering information in order to determine the relationships between a person's problem behavior and aspects of their environment [PBS]. Once the relationship between a person's problem behavior is determined, it is possible to identify specific events that predict and maintain behavior and design a support plan that effectively addresses those variables [PBS]. The first thing to do in conducting functional behavioral assessment is to identify and define the problem behavior in broad terms. The second step is to identify the relevant dimensions of the problem behavior and define them in an explicit and unique way. In the third step, information should be collected from various sources (questionnaires, interviews, direct observation) on: times when the behavior does/does not occur, location of the behavior, conditions when the behavior does /does not occur, individuals present when the problem behavior is most/least likely to occur, events or conditions that typically occur before the behavior, events or conditions that typically occur after the behavior, common setting events (e.g. weather condition), and other behaviors that are associated with the problem behavior). The collected information is analyzed and based on information on the social/environmental context, antecedent and consequent events and past events; when, where, with whom, and under what conditions certain behavior is most /least likely to occur is predicted. In fourth step, comparison and analysis of all the collected information with the help of analysis techniques such as data triangulation and problem pathway analysis is conducted to determine which specific social, affective, and environmental conditions are associated with the problem behavior. In the end, a hypothesis statement that explains why the problem behavior occurs(what the functions or purpose of the behavior) is developed and tested. The hypothesis statement can then be used to predict the social/environmental conditions (the context) within which the behavior is most

likely to occur. The known functions of the behavior help in establishing an effective behavioral correcting strategy. [CECP 98]

3. **Evaluating Behavioral Correcting Strategy:** evaluating the effectiveness of behavioral correcting strategy is also referred to be functional analysis of behavior. It includes first systematic observation of the problem behavior to obtain a response frequency baseline and then further observation after the behavioral correcting strategy is implemented to record any changes in behavior that may occur.

There are three behavioral observation approaches for collecting behavioral data: self report, direct behavioral observations, and psychophysiological recordings [Anthony & Karen 86]. The usefulness of these three approaches depends on what the goal or purpose of the assessment is and what kind of response is to be collected and assessed.

3.2.1 Self Report

Self report approach includes data collected in behavioral interviews, in questionnaires and inventories, and through self-monitoring procedures. Although information about all three response modes (i.e overt-motor, physiological-emotional, verbal-cognitive) can be collected by means of self report, it is used primarily to assess the verbal-cognitive response system. [Anthony & Karen 86]

The first method of self report, the behavioral interview, is probably the least structured assessment strategy. One reason is the difficulty in operationalizing, measuring, manipulating, and controlling the complex variables inherent in interviewing. Much has been written and considerable research has been conducted on aspects of an interview. However, there is no consensus as to how to conduct an interview. The only areas in which there is consensus is as to what should be included in an interview. Three steps should be conducted in every interview: specifying the problem (identification of behaviors and classifying them in to behavioral excesses, deficits, and assets), antecedent analysis, and consequence analysis. [Anthony & Karen 86]

The second major method for collecting self-report data is through some written behavioral questionnaire or inventory. Aside from being more objective and standardized than an interview, there are several other advantages to questionnaires and inventories. Wolff and Merrens (1974) indicated that written questionnaires are quick, inexpensive, easy to administer, and can provide objective measures for various assessment purposes. Multiple response modes such as physiological, motoric, and cognitive behaviors can be gathered using questionnaires. Despite many advantages of questionnaires, some limitations exist. One of the primary limitations concerns the indirect nature of the data collected. Questionnaire responses (i.e. pencil marks) are far removed from the actual behaviors. Without evidence to suggest a significant relationship between questionnaire

responses and more direct target behaviors, the data obtained from questionnaires will have limited utility. [Anthony & Karen 86]

The third method of self report requires participants to monitor various aspects of their behavior in daily living situations. This method is frequently required if a private or covert behavior, which can not be observed by anyone other than the participant, is to be recorded. By definition self-monitoring is the act of systematically observing and recording aspects of one's own behavior. There are three major kinds of self-monitoring procedures : frequency related measures, times-related measures, and outcome measures. Selection of a particular recording procedure depends upon the amount of information it conveys, how accurately that information can be obtained, and the ease with which the behaviors are recorded. [Anthony & Karen 86]

- A. Frequency-Related Measures: the frequency count or event-sampling method is the most common self-observation procedure currently in use. It is appropriate primarily when the behavior to be observed occurs as a discrete event with an identifiable beginning and end. Behaviors, which are of relatively short duration and frequency, are most amenable to this type of recording. Event recording keeps track of each occurrence of the behavior of interest. The behavior is recorded, the time of the occurrence is noted, and a response rate may be calculated. In order to obtain an accurate assessment of frequency, low rate behaviors must be recorded over a longer period of time than high rate behaviors. In addition collecting frequency data for baseline informational purposes, event recording can be collected along with discriminative stimuli data (helpful in identifying controlling antecedents and consequences) and additional information of interest such as date and time. One disadvantage of event recording is that discrete events destroy the continuity of behavior.
- B. Time-Related Measures: time sampling is used primarily when high rate behaviors are to be recorded, and /or when it is difficult to determine the discrete occurrence of a particular behavior. The most common form of time-sampling method presently employed in behavioral research is that of interval recording. Here, the individual is asked to determine whether or not a particular target behavior has occurred during a series of repeated brief time periods. The observational period is the total time wherein the observer monitors and records the target behavior. This observational period is divided into the observational interval (during which the observer is directly viewing the behavior) and the recording interval (wherein the observer records what he or she has monitored). When high frequency behaviors are recorded, an increase in sampling occurs and it may become more difficult for the client to accurately self-monitor. In such instances, an external prompt for observing may be provided. Such a cueing device can be mechanical or electronic and should signal the observer to begin monitoring the target behavior. Intervals can be observed and scored in terms of how frequently the target behavior occurs. When behavioral intensity is important a rating scale can be used. In general, interval sampling should be relatively continuous so as to build the most accurate representation of the behavior of

interest. Duration methods are yet another form of time-related measures. These methods are used to record the amount of time an individual exhibits the behavior of interest. Duration measures should be used if the behavior is not discrete, if it varies in duration, or if it occurs for long period of time.

- C. Outcome Measures: the occurrence of behavior can be measured also as a correlated outcome of some other behavior (e.g. measuring bald spots to determine hair pulling behavior).

3.2.2 Direct Behavioral Observation

Direct recording of an individual's behavior by some independent observer is probably the hallmark of behavioral observation. This method is used primarily in the assessment of the overt-motor response mode by having the client's behavior directly observed by trained recorders or other individuals. Direct observations can be extremely useful in the specification of the target behavior and its controlling events as well as in the evaluation of behavioral measures. Behavior can be observed in naturalistic settings or in contrived (analogue) settings in a laboratory. [Anthony & Karen 86]

Numerous considerations influence the selection and design of any direct behavioral observation system. Among the most important of these are the purpose for making observations, the subject matter of the observations, specific characteristics of that subject matter, the amount of behavior to be observed, the minimum information needed to establish the quality of the data to be produced, and available resources. [Anthony & Karen 86]

Following the decision to use direct observational procedures, it is frequently necessary to design a system from scratch to capture the phenomena of interest. The goal of this process is to design a system that produces data suitable for the purposes for which it was designed and that are accurate and replicable. At this point, major issues are defining and quantifying behavior, selecting a recording procedure, making sampling decisions regarding how much data to collect and whether (and how) to restrict the situations in which data are collected. [Anthony & Karen 86]

Defining behavior for observation involves creating written guidelines for observers such that after training they extract and record relevant behavior from ongoing activity and exclude from their records behavior that fails to fit guidelines [Anthony & Karen 86]. To define and quantify behaviors first a behavioral taxonomy, which is a classification framework, has to be designed [7]. Behavioral taxonomy is representation of an individual's or group's behavioral repertoire, containing a list of behaviors along with an operational definition of each behavior. Each behavior is seen as a category that is defined by the different attributes that make up the behavior. These categories group behaviors according to some commonality of function or the outcome of a behavior, such as feeding, grooming, communicating, etc. The definition of each behavior should be precise, valid, reliable and objective. Some behaviors may contain other behaviors, but

the included behaviors must be defined also, for example: the definition of the behavior following, may include the behavior watching. [7]

The categories or clusters of behaviors should be exhaustive (completeness, that contain all possible behaviors), mutually exclusive (behaviors must fit into one and only one category), and based on operational definition of behavior. [7]

Numerous procedures are available for recording behavior once observational categories have been established. They are described as follows. [Anthony & Karen 86]

- A. Real-time (Continuous) Recording Procedures: they provide ongoing records of behavior as they transpire. Observers ordinarily utilize equipment to indicate when each behavior in the observation code begins and ends, and the system automatically records elapsed time. This kind of system permits derivation of frequency, duration and related response properties, rates, and sequence of behaviors included in the system. Other recording procedures are frequency (event) recording, in which each instance of a target behavior is recorded as it occurs, and duration recording, which records the time between onset and offset of the behavior.
- B. Momentary Time Sampling: with this kind of recording procedure, the observer is cued to record a target subject's behavior at specific points in time. This recording procedure is also called instantaneous sampling. It is investigated by several researchers that momentarily time-sampling procedures distort observational data when compared with data yielded by continuous assessment procedures.
- C. Interval Recording: in this recording procedure the observation session is divided into time blocks, usually of equal duration. Although observers can employ frequency and/or duration recording within intervals, by far the more usual interval recording procedure involves noting whether the behavior occurred at all or during a specified percent of the interval (partial-interval sampling), throughout the entire interval (whole-interval sampling), or at a specific moment during the interval, such as at the beginning, the end, or at the sound of a timer. It is proved that interval recording procedures, both whole and partial, often provide biased estimates of actual behavior.

After a direct observational system has been designed, refined, shown to be accurate, and observers have been trained to use it, data collection can commence. During data collection, several factors can influence data quality. Among these are observer expectations, reactivity of subjects to observation, interactions between experimenter and observers, and observer drift (a gradual change in the use of category definitions). It has been shown that informing observers of experimental hypothesis or variables under investigation lead them to observe and record data in biased fashion. One suggestion to mitigate expectation bias is to focus on operational definition of specific behaviors. The

gradual change in use of category definitions can be minimized if there is consistent observer recalibration during data collection. [Anthony & Karen 86]

3.2.3 PsychoPhysiological Recordings

Direct measurement of psychophysiological response system (e.g. cardio vascular responses) is used for behavioral assessment in the physiology-emotional response mode. There are numerous ways to measure physiological responses. Each way is determined by the physical characteristics of the biological response and the available technology. Brown (1972) describes two basic types of physiological responses: bio-electric potentials and responses derived from physical characteristics of the organism. The minimal instrumentation necessary for measuring physiological responses consists an input transducer (or electrode), an amplifier, and an oscillograph (pen writer). [Anthony & Karen 86]

Recent evidence from psychophysiology and personality theory suggest that behavior is situation-specific. The concept of situational specificity implies that behaviors occur in a given environmental context and do not represent an underlying biological or personality trait. Therefore, the assessment of physiological responses is valid only in relation to a specific set of stimulus events. [Anthony & Karen 86]

Although the expense involved in psychophysiological recording is great when compared with other behavioral assessment methodologies, there is justification for its use. The measurement of physiological reactivity to environmental events is the only objective means of assessing a covert event. [Anthony & Karen 86]

3.3 Computer Based Behavioral Observation

It is advantageous to use instrumentation during behavioral observation since the sensory abilities of human being allow observation of only a small fraction of the stimuli or energy in our environment. So far numerous instrumentation devices have been developed that work for specific behavioral responses [Anthony & Karen 86]. In this section, the role of computer system in conducting behavioral observation is only emphasized.

Computers can be used in collecting self-report information and conducting questionnaire assessment. The advantages of computer assisted questionnaire assessment include savings in professional time, reduced error variance through standardized administration procedures, and ease of data collection and analysis. However, there are several important issues, which need to be examined in order to increase the utility of this assessment approach. Such considerations as the validity of information obtained via computer-administrated questionnaires, appropriateness of the approach for various subgroups of

clients, and client acceptability and computer comfort must be addressed. With proper attention to these concern, computer assisted questionnaire assessment has the potential to become a valuable method for diagnosis and screening, treatment planning, and outcome evaluation. [Anthony & Karen 86]

Beside the role of computers in conducting questionnaire assessment, they are used also in supporting collection and assessment of psychophysiological responses , and direct observation behavioral assessment. There are software systems that support the process of behavioral observation studies from the initial collection of psychophysiological responses or over-motor responses till the analysis of the collected behavioral data and presentation of output results. The most known of such software system are The Observer Software developed by the Noldus Information Technology and CAMERA, a system for behavioral observation from video recordings. A detailed explanation of these two software systems is given in the following sections.

3.3.1 The Observer

The Observer is a professional tool for the collection, analysis, presentation and management of observational data. It allows to score data quickly and accurately, whether these involve activities, postures, movements, positions, facial expressions, social interactions, or any other aspect of behavior. The Observer supports the entire research process, from the design of observational studies and coding schemes, through collection of behavioral data to the analysis and presentation of results. [Noldus 03]

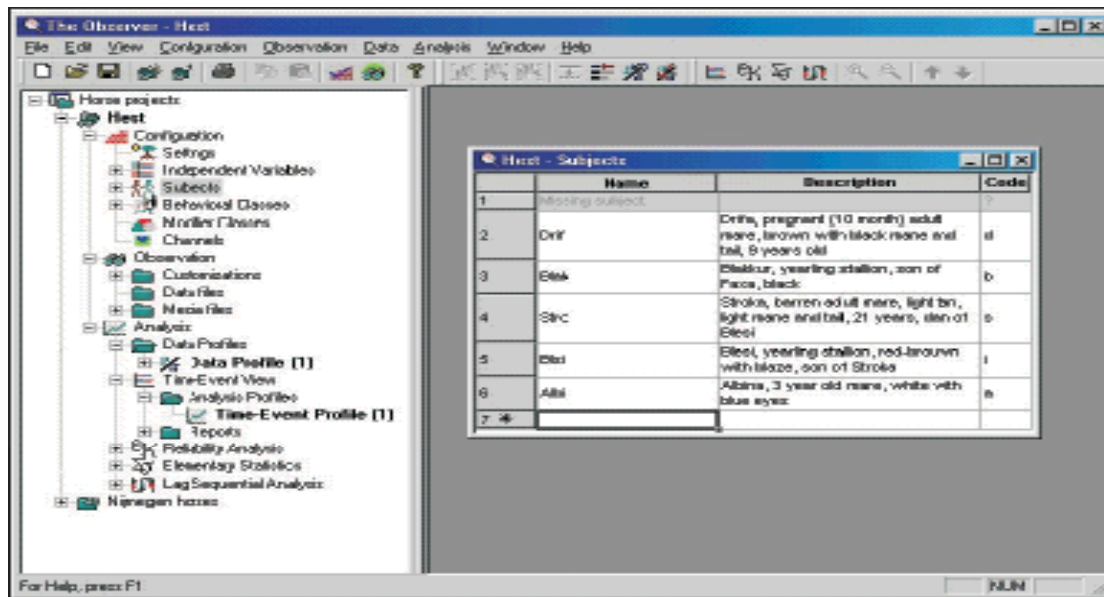
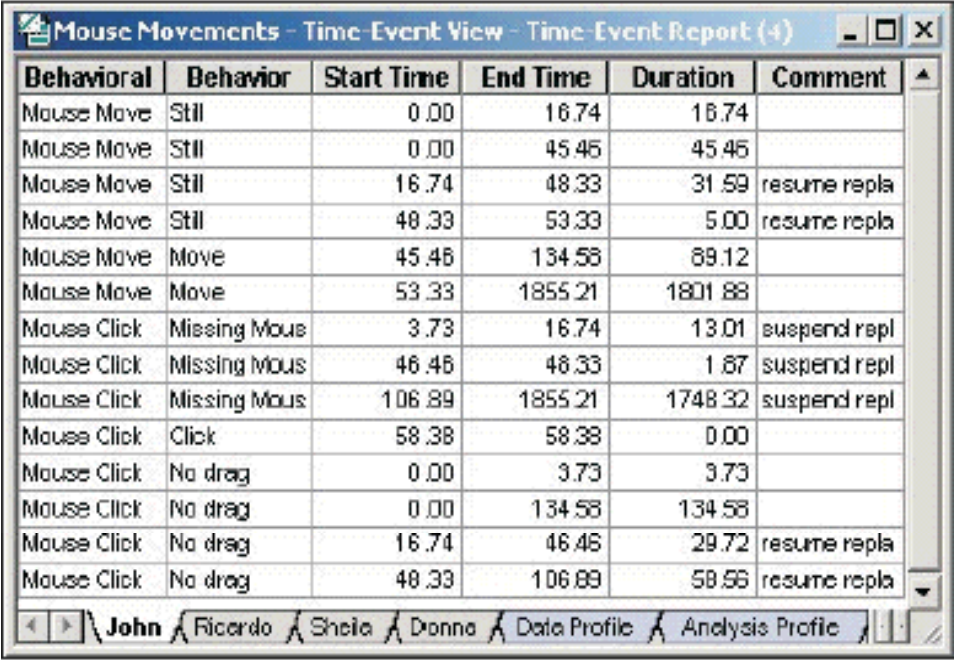


Figure 3.1 The main module of The Observer Software

The Observer Software is provided in three modules: The Observer Basic, The Observer Mobile, and The Observer Video-Pro. The Observer Basic comprises everything to carry out live observations from a fixed position. It contains all the software necessary to set up coding schemes, collect observational data, manage projects and analyze collected observational data. The Observer Mobile is used to collect data on the move using a handheld computer as event recorder. It is used when observing very active subjects in large premises, such as hospitals or factories, or when video recording is not permitted. The Observer Video-Pro combines the power and flexibility of the Observer Basic with the convenience of video and multimedia technology. It is valuable tool when time-accuracy is of particular importance, or when there is a need to observe the behavior of several individuals at the same time. [Noldus 03]

The Observer provides powerful analysis and presentation options accessible by mouse clicks and keystrokes. The three different way of presentation tools are described briefly as follows: [Noldus 03]

1. Video Play List: an easy-to-use tool for compilation of specific video episodes. It creates a list of video clips for an immediate presentation on-screen or via a computer projector.
2. Time-Event Table: it shows a chronological listing of all recorded events, sorted into columns based on subjects and classes of behavior and visualizes changes in state and occurrence of events in a course of time.



Behavioral	Behavior	Start Time	End Time	Duration	Comment
Mouse Move	Still	0.00	16.74	16.74	
Mouse Move	Still	0.00	45.45	45.45	
Mouse Move	Still	16.74	48.33	31.59	resume repla
Mouse Move	Still	48.33	53.33	5.00	resume repla
Mouse Move	Move	45.45	134.58	89.12	
Mouse Move	Move	53.33	1855.21	1801.88	
Mouse Click	Missing Mous	3.73	16.74	13.01	suspend repl
Mouse Click	Missing Mous	46.45	48.33	1.87	suspend repl
Mouse Click	Missing Mous	106.89	1855.21	1748.32	suspend repl
Mouse Click	Click	58.38	58.38	0.00	
Mouse Click	No drag	0.00	3.73	3.73	
Mouse Click	No drag	0.00	134.58	134.58	
Mouse Click	No drag	16.74	46.45	29.72	resume repla
Mouse Click	No drag	48.33	106.89	58.56	resume repla

Navigation: < > \ John / Ricardo / Sheila / Donna / Data Profile / Analysis Profile

Figure 3.2 The Time-Event Report for the Time-Event Table

3. Time-Event Plot: provides a graphical representation of observational data and show a quick overview of what happened and when.

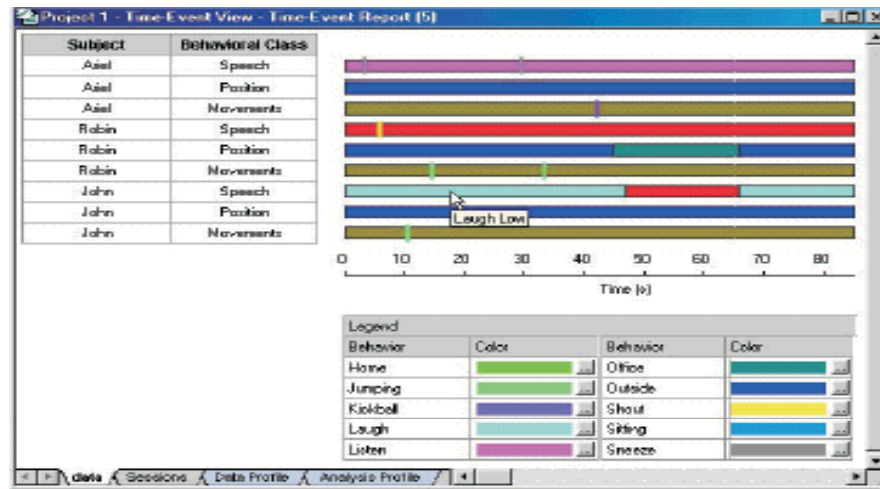


Figure 3.3 The Time-Event Report for the Time-Event Plot.

The three analysis tools are described briefly as follows: [Noldus 03]

1. Elementary Statistics: this analysis tool calculates crucial descriptive statistics on the frequency and duration of events or states; it also calculates statistics for subjects, behaviors and modifiers, in any combination; and categories can be split or lumped using nesting levels to create complex queries, define iterative time windows, and average data across observations.

The screenshot displays a statistics report table. The columns represent behaviors: Talk, Laugh, Listen, Standing, Walking, Running, and Sitting. The rows represent statistical measures for Session 1 and Session 2. The data is sorted by session, then by behavior, and then by subject.

		Talk	Laugh	Listen	Standing	Walking	Running	Sitting
	Behavioral class	Speech	Speech	Speech	Movements	Movements	Movements	Movements
Session 1	Total number	2	2	0	1	1	0	0
Session 1	Rate	1.77	1.77	0.00	0.89	0.89	0.00	0.00
Session 1	Total duration	50.22	17.40		45.20	22.42		
Session 1	Total duration (% of observation)	59.57	20.64		53.61	26.59		
Session 1	Mean duration	25.11	8.70		45.20	22.42		
Session 2	Total number	1	0	2	2	1	1	1
Session 2	Rate	0.71	0.00	1.42	1.42	0.71	0.71	0.71
Session 2	Total duration	10.28		74.03	38.06	17.23	27.79	2.33
Session 2	Total duration (% of observation)	12.19		87.81	43.64	20.44	32.96	2.78
Session 2	Mean duration	10.28		37.02	18.48	17.23	27.79	2.33

Figure 3.4 The Statistics Report. The Report is sorted with the statistics in the rows, behaviors in the columns and each subject on its own sheet.

2. Reliability Analysis: it measures the level of agreement between pairs of data collectors. The reliability result is reported as index based on the degree of agreement between two data files. Each result is scored as a match or an error in regards to timing and behavioral information. Tolerance for discrepancy is controlled by a user defined time window. The reliability report consists of three different outputs, each of which is placed on its own view in the reliability report window: statistical measures, confusion matrix, and list of comparisons.
3. Lag Sequential Analysis: is a collection of techniques developed for the study of the temporal structure of a sequence of events. It technique determines how often certain events precede or follow other events. It generates matrices, which can be modified to include or exclude specific subjects, behaviors or modifiers. It can include also state lags or time lags of user-defined length. For advanced matrix analysis, output can be exported to MatManTM.

		Subject 1	Subject 1	Subject 1	Subject 1	Subject 1	Y0
		Look at shelf	Take product	Read label	Product back	Buy product	
Subject 1	Look at shelf	0	4	0	1	0	3
Subject 1	Take product	0	0	4	2	1	0
Subject 1	Read label	1	0	0	3	0	1
Subject 1	Product back	2	1	0	0	0	2
Subject 1	Buy product	2	0	0	0	0	0
XD		3	2	0	0	1	-

Figure 3.5 The Lag Sequential Report.

Behavioral data observed by The Observer can be also analyzed further using a Theme, a unique tool for detection and analysis of hidden patterns in behavior. [Noldus 03]

The Observer, due to its flexible and generic design, can be used in a broad variety of research fields that approach behavior in very different ways. The most applicable research fields are: [Noldus 03]

- Animal behavior: ethology, behavioral ecology, entomology, animal husbandry, zoo biology and conservation.
- Neuroscience: preclinical CNS drug research, brain-behavior relationships, behavioral disorders, and pharmacology.
- Toxicology: screening chemical substances for behavioral side effects.

- Psychology: behavioral development, parent-child interactions, aging, communication, education, social interactions, language acquisition, cognition, and psychological assessment.
- Medicine: pediatric, neurology, psychiatry, rehabilitation, etc.
- Ergonomics: usability testing of software user interfaces or hardware designs, working posture analysis, workplace design, physical workload, occupational health and safety.
- Human factors studies: task analysis, labor studies, analyzing human-computer interaction, evaluating man-machine interfaces, traffic research, driving, and flight simulator studies.
- Sports observations: notational analysis, game and tactics analysis, evaluation during scouting, posture and movement studies, and feedback and review during training.
- Training, consumer behavior and advertising research, time studies, efficiency research, legal trial analysis, industrial engineering, and etc.

3.3.2 CAMERA

CAMERA is a system for collecting and correcting ethological data. Its hardware and software enable the easy registration and encoding of complex behavioral interactions from video recordings. The accurate and comprehensive description of complex behavior is a major goal in ethological studies of behavior. In behavioral research, video recordings are often used to meet this goal, to make it possible to examine behavior repeatedly. Yet the processing of directly-observable live behavior by the coding of video-recorded behavior into sequences of distinct events is prone to error and distortion. These errors originate from two sources: the observer and the coding instrument. How well the observer performs depends on the number of different events, the clarity of the definitions, the rate of the occurrence of the events, the ergonomics of the technical task, and on vigilance and training. CAMERA has been designed to reduce coding errors from both sources. [6]

A key feature of CAMERA is the writing of an invisible time code in each video frame during recording. This code is automatically retrieved and checked during replay of the video recording. Every time an event is coded by pressing a corresponding button on the event keyboard, the start and end time of the event, as well as an event label defined by the researcher, is stored in the memory of the personal computer. Encoded behavioral records can be graphically displayed and inspected in interactive routines as bout sequence plots, log survivor plots and cumulative duration plots. Summary statistics such as means and standard errors of duration and frequency of each behavioral item, are also available. Data can be output in ASCII form to allow analysis by standardized statistical packages. [6]

Chapter Four: User Behavior Observation Methods in Interactive Computer Applications

4.1 Introduction

Interactive computer systems are systems where end-users (human beings) are frequently interacting with the computer system to accomplish a certain task and include multi-modal flow of information such as text, graphics, audio, and video. The emphasis in this chapter is to discuss available concepts and techniques for recording and assessing end-users behavior while they are interacting with an interactive computer system.

The behavior of end-users of interactive computer applications is usually collected and assessed for usability evaluation and user modeling. User modeling, which is an activity that attempts to describe the user's behavioral patterns, preferences, interests, and characteristics, can be conducted for providing personalized or adaptable web applications. Personalized web applications are those applications that are responsive to the individual needs of each user. [9] [Geo& Michael&Daniel 00]

Several research works have been conducted regarding usability evaluation of interactive computer applications. A detailed discussion on usability evaluation is presented in the following section.

4.2 Usability Evaluation Methods

Much of the success of interactive applications such as web application depends on how easy it is for users to use the application; users must be able to accomplish their intended tasks quickly and without error. Combining design principles and usability testing enhances the ability of interactive applications to meet these goals. [Craig]

The emphasis in this report is not how to design usable interactive applications but how to collect and evaluate real user behaviors for the purpose of identifying usability problems with interactive applications. The reason behind understanding how real user behaviors are collected and evaluated for the purpose of usability testing is to gain knowledge of how user behavior can be collected and evaluated in conducting usability evaluation in the domain of Virtual Environments. To conduct any method of usability evaluation, the following main steps have to be conducted: [10]

- Determining a purpose
- Identifying test objectives
- Designing the tests
- Running the experiment

Several usability evaluation methods are available. The most known are Direct Observation, Interview, Questionnaires, Focus Groups, Heuristic Evaluation, Cognitive

Walkthrough, Layout Analysis, Performance Measuring, and Logging User Actions. User behaviors has to be collected and assessed in order to conduct usability evaluation methods except in some methods such as Heuristic Evaluation, Cognitive Walkthrough, and Layout Analysis which are done by experts on the field with out the involvement of users. The main goal of the assessment of the collected user behaviors through various methods is to evaluate if an interactive application under study is usable or not. A usable interactive applications has the following characteristics: lets the user execute tasks quickly and with minimal errors; has a small or no learning curve; performs well in multiple environments; performs well for different users; leverages the user's existing skill set; and consistently enables the user to transfer his existing skills. [Craig]

4.2.1 Direct Observation

Direct observation is one of the usability testing methods where observers observe representative users while performing a scenario of tasks and record the actions performed by video or hand. Direct observation can be conducted in different ways: field observation, thinking-aloud, and co-discovery learning. [10]

In Observation field technique, users are observed in the field. It is often the best way to determine usability requirements. Traditional usability testing, while providing a laboratory environment that makes data collection and recording easy, also removes the user and the product from the context of the workplace. Sometimes, it's best to see exactly how things are done in the real world. Part of field observation is inquiry; that is, interviewing users about their jobs and the ways they use the product to be tested. The other part is observation, which is watching people use the product in the way they normally would in the course of day-to-day life. Observation field technique is started by arranging for a field visit with users and choosing a variety of users of the product to be tested from different workplaces, industries, and backgrounds. This technique is best used in the early stages of development, when you need to know more about the issues surrounding the use of a product rather than actual metrics. [10]

Thinking Aloud protocol is a popular technique used during usability testing. During the course of a test, where the participant is performing a task as part of a user scenario, you ask the participant to vocalize his or her thoughts, feelings, and opinions while interacting with the product. Thinking aloud allows you to understand how the user approaches the interface and what considerations the user keeps in mind when using the interface. If the user expresses that the sequence of steps dictated by the product to accomplish their task goal is different from what they expected, perhaps the interface is convoluted. Although the main benefit of the thinking aloud protocol is a better understanding of the user's mental model and interaction with the product, you can gain other benefits as well. For example, the terminology the user uses to express an idea or function should be incorporated into the product design or at least its documentation. Thinking aloud is a cheap way of getting a lot of good qualitative feedback during testing. [10]

Co-discovery is a type of usability testing where two participants attempt to perform tasks together while being observed. The advantage of this method over the thinking aloud protocol is two-fold: in the workplace, most people have someone else available for help; and the interaction between the two participants can bring out more insights than a single participant vocalizing his or her thoughts. This technique can be used during any phase of development. It is ideal for Computer-Supported Cooperative Work (CSCW) products, groupware, and other products designed to be used by workers in team environments. [10]

4.2.2 Interview and Focus Group

Interviews and focus groups are conducted to query users about their experiences and preferences with the product to be tested by raising issues and let the users discuss about them [10]. The methods of interviewing include unstructured interviewing and structured interviewing. Unstructured interviewing methods are used during the earlier stages of usability evaluation. The objective of the investigator at this stage is to gather as much information as possible concerning the user's experience. The interviewer does not have a well-defined agenda and is not concerned with any specific aspects of the system. The primary objective is to obtain information on procedures adopted by users and on their expectations of the system. Structured interviewing has a specific, predetermined agenda with specific questions to guide and direct the interview. Structured interviewing is more of an interrogation than unstructured interviewing, which is closer to a conversation. [11]

In Focus Group method, about 6 to 9 users are brought together to discuss issues relating to the system. A human factors engineer play the role of a moderator, who needs to prepare the list of issues to be discussed beforehand and seek to gather the needed information from the discussion. This can capture spontaneous user reactions and ideas that evolve in the dynamic group process. [11]

Interview and Focus Group techniques can be used at any stage of development, depending on the questions that are asked. Often, interviews or focus groups are used after products are shipped to assess customer satisfaction with the product. More likely, though, interviews and focus groups are held at very early stages of development, when the product requirements are still not firm. Focus groups are then held to extract user requirements prior to initial design. [10]

4.2.3 Questionnaire

A questionnaire is a method for the elicitation, recording, and collecting of information. It is conducted by distributing questionnaires, which are made up of items to which the users (respondents) supply answers or reactions. There are three basic types of questions: factual-type questionnaires, opinion-type questions, and attitude questions. [10]

Factual-type questions ask about public, observable information that it would be tedious or inconvenient to get by any other way. One example is data about users such as age, experience, and expectations. Opinion-type questions ask the respondent what they think about something or someone. An opinion survey is concerned with finding out how popular someone or something is. Opinion questions direct the thought of the respondent outwards, towards people or artifacts in the world out there. Responses to opinion questions can be checked against actual behavior of people, usually, in retrospect. Attitude questions focus the respondent's attention to inside themselves, to their internal response to events and situations in their lives. Examples are questionnaires that ask the respondent what their attitudes are to working with a particular product. These are generally called satisfaction questionnaires. [10]

Several standard usability evaluation questionnaires have been designed. The best known standard questionnaire, which has been developed, validated and standardized in an international basis, is SUMI (Software Usability Measurement Inventory). SUMI consists 50 items, in which each subscale was represented by 10 different items. The response categories are agree, do not know, and disagree. The subscales are grouped into five: efficiency, affect, helpfulness, control, and learnability. Efficiency measures the degree to which users feel that the software assists them in their work and is related to the concept of transparency. The Affect subscale measures the user's general emotional reaction to the software, it may be glossed as Likeability. Helpfulness measures the degree to which the software is self-explanatory, as well as more specific things such as the adequacy of help facilities and documentation. The Control dimensions measures the extent to which the user feels in control of the software, as opposed to being controlled by the software, when carrying out the task. Learnability, finally, measures the speed and facility with which the user feels that they have been able to master the system, or to learn how to use new features when necessary. [11]

One advantage of questionnaires is that a usability questionnaire gives feedback from the point of view of the user. If the questionnaire is reliable and is used according to the instructions, then the feedback is a trustworthy sample of what can be collected from the whole user population. Another advantage is that measures gained from a questionnaire are to a large extent, independent of the system, users, or tasks to which the questionnaire was applied and can be used to compare the perceived usability of a word processor with an electronic mailing system and the ease of use of a database as seen by a novice and an expert user. Additional advantages are that questionnaires are usually quick and therefore cost effective to administer and to score. [10]

The biggest disadvantage is that a questionnaire tells you only the user's reaction as the user perceives the situation. Thus some kinds of questions, for instance, to do with time measurement or frequency of event occurrence, are not usually reliably answered in questionnaires. A questionnaire is usually designed to fit a number of different situations. Thus a questionnaire cannot tell in detail what is going right or wrong with the application you are testing. [10]

4.2.4 Performance Measuring

Performance measuring method is used to obtain hard, quantitative data from users actions while they are performing a scenario of tasks. Performance measuring tests are done usually in usability laboratory. Examples of measurements that can be obtained by performance measuring are: [12]

- The time users take to complete a specific task.
- The number of task of various kinds that can be competed within a given time limit.
- The ratio between successful interactions and errors.
- The number of user errors.
- The number of commands or other features that were never used by the user.
- The frequency of use of the manuals and/or the help system, and the time spent using them.
- The proportion of users using efficient working strategies in case there are multiple ways of performing the tasks.

Performance measurement is used in initial stages of design to provide benchmarks for the design process. It's also used during the design cycle to measure the work done thus far against those benchmarks. [10]

4.2.5 Logging User Actions

Logging involves having the computer automatically collect statistics about the detailed use of a system. It is useful because it shows how users perform their actual work and because it is easy to automatically collect data from a large number of users working under different circumstances. Typically, an interface log will contain statistics about the frequency with which each user has used each feature in the program and the frequency with which various events of interest (such as error messages) have occurred. Statistics showing the frequency of use of commands and other system features can be used to optimize frequently used features and to identify the features that are rarely used or not used. Statistics showing the frequency of various error situations and the use of online help can be used to improve the usability of future releases of the system by redesigning the features causing the most errors and most access for online help. This technique can be used at the test or deployment stages of software development. [14]

Logging is usually achieved either by instrumenting low-level parts of the system software, such as keyboard and mouse drivers, or by modifying the software of interest. The latter is much preferred, since it makes it easier to log events of interest. If the only available data is raw input and output, it becomes much harder to analyze the higher-level events of interest for system usability, such as feature use or error situations. If the instrumented system runs on a mainframe or on workstations with a shared file space, it is easy to collect logging data by simply copying the log files from each user at regular intervals. Otherwise, it may be necessary to collect log data through electronic mail --

either automatically or by asking the users to periodically run a small script that sends off the log file. Or, the logs may be collected by going around to the users and copying floppies. In addition to statistical use of logging data, it is also possible to log complete transcripts of user sessions either for use in later playback or for analysis of patterns of use, such as what commands are issued next after an error situation. A final use of logging data is to study the user's detailed use of a user interface to find usability problems that may not be apparent when observing users. The logging data can vividly show the statistics of one action (e.g. clicking on an object) from many users by integrating the logging data with the user interface. Since the logging data only shows what the users did but not why they did it. It would be better to combine logging with other methods such as interviews, where users are shown data about their own use of the system and asked to explain their activities. [14]

Tools have been designed and developed not only for automatically logging the detailed use of a system by users but also to detect usability problems by performing automatic evaluation. One of such tools is WebRemUSINE. WebRemUSINE is an environment designed and developed at the HCI Group of ISTI-C.N.R. in order to perform intelligent analysis of Web browser logs using the information contained in the task model of the application. WebRemUSINE performs an automatic evaluation of a Web site providing the evaluator with a set of measures, concerning also group of users, useful to identify usability problems derived from a lack of correspondence between how users perform tasks and the system task model. [3]

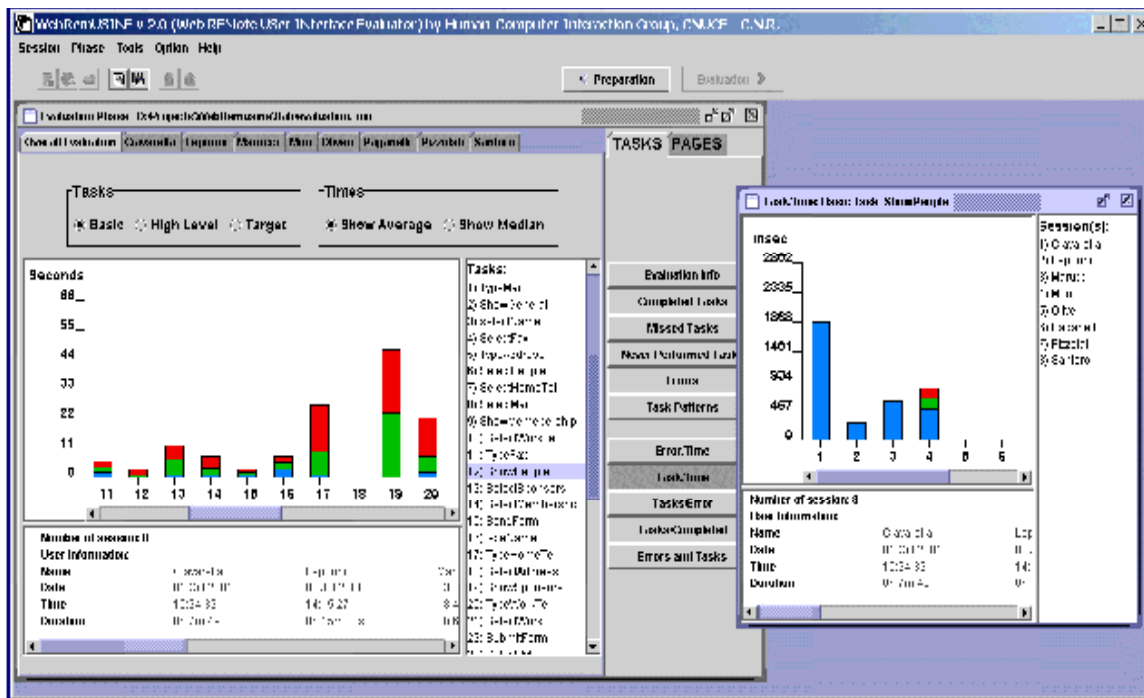


Figure 4.1 Screenshot of the WebRemUSINE tool.

The inputs for the tool are the task model and the log files recorded during the test sessions. The environment is mainly composed of three modules: the ConcurTaskTrees editor specify task models in a hierarchical structure enriched with a number of flexible temporal relationships among such tasks (concurrency, enabling, disabling, suspend-resume, order-independence, optionality, ...); the browser logging tool that has been implemented to record user interactions and that can be easily installed in a Web site; WebRemUSINE, the java tool able to perform an analysis of the files generated by the logging tool using the task model created with the CTTE tool. This approach supports remote usability evaluation of Web sites. The evaluation performed provides information concerning both tasks and Web pages. These results allow the evaluator to analyze the usability of the Web site from both viewpoints, for example comparing the time to perform a task with that for loading the pages involved in such a performance. WebRemUSINE also shows an analysis of each log files. [4]

Chapter Five: Discussion on Insights Gained from Related Domains

In this chapter whether and how the knowledge gained from other domains: real environment and interactive computer applications can be applied in conducting behavioral observation in Virtual Environments is discussed.

5.1 Insights from Behavioral Observation Methods in Real Environments

It has been explained earlier how much user interactions or activities in Virtual Environments is closely related to interactions and activities in real environments. This fact triggers the idea of attempting to apply the concepts and approaches of behavioral observation, which are discussed in detail in chapter three, in observing and assessing the behavior of end-users or other autonomous entities in Virtual Environments. In this section, we first try to define behaviors in Virtual Environments based on the terms and concepts used in defining and representing human behavior in real environments. Then, comparison and evaluation of all the approaches we discussed so far is given and are assessed how to apply them in collecting and assessing behaviors in Virtual Environments.

The definition and scope of human behavior in real environment can be applied also to define and categorize behaviors that occur in Virtual Environments. All Virtual Environment applications have at least one End-user who interacts with the virtual environment via the input and output modalities supported by the computer system that runs the Virtual Environment application. Usually a 3D representation of the end-user(s) or other object like a vehicle in the case of Driving Simulator or a pointer icon is displayed in the virtual environment where he/she interacts. The status, position, activity, movement, posture, and etc. of the 3D representation are changed when the end-user(s) manipulates the input devices provided by the computer system that runs the Virtual Environment application.

It has been noticed from the discussion on the overview of Virtual Reality technology that there is no a direct transformation of what the end-user do and what the 3D representation do. For example, the end-user need not to run actually if he/she wants to run in the Virtual Environment. The 3D representation can run when the end-user presses a button in a certain sequence. Other example is when the end-user wants to manipulate objects. In real environment, we first grab the object and then do whatever manipulation we want to do like changing size or moving it to other location. If the object is far away from us we walk close to the object and then grab the object. However, these activities can not be directly transferred to the 3D representation since the end-user is physically constrained in a confined space. One technique for grabbing or selecting far away objects is to use laser pointing, in which a ray is extended from the user's hand into the Virtual Environment to intersect the object we want to select.. The way activities done by the end-users and how these activities are reflected in the Virtual Environment depends on

the kind of interaction technique used by the Virtual Environment application under consideration.

Human beings in real environment can do any kind of activities like kicking and grabbing; can do movements like running, walking, jumping, and climbing; can change position or posture; can do gesturing; can talk or cry or scream; and show different facial expressions like angry, happy, and sad. And they sense their environment via the five kind of sensory channels: vision, audio, touch, taste, smell. In the Virtual Environment case, what can be sensed or perceived by the end-user is constrained by the available output modalities. Currently, visual signals, audio signals, and touch signals can be sensed by end-users of most Virtual Environment applications. If the end-user of a Virtual Environment application wants to do any kind of response like kicking or grabbing a virtual object or walking in the virtual world, he/she has to use a certain kind of interaction technique. There are two approaches to implement interaction techniques: direct and indirect approach. In the case of direct interaction techniques, the end-user has to do similar to the way he/she does things in real environment. For example, to move a virtual object from one location to another, the end-user has to move his hand toward the virtual object then touch the object and move his hand toward the target location. In the case of indirect interaction technique approach, the end-user has to extend ray toward the virtual object and presses some button to select it and then point the ray toward the target location and presses a button to put it in that location. Even though direct interaction technique is comfortable and easy to use them, it is difficult to implement direct interaction technique for some activities, for example running or flying. For such kind of activities, an indirect interaction technique is preferable.

From the above discussion, one thing can be noticed, all the responses or activities done by the end-users have no often the same meaning or purpose to the responses or activities in the virtual world. For example, what we see an end-user riding a stationary bike-like input device might turn out to be walking activities in the virtual world. This fact is very important issue in designing observation method in Virtual Environments.

The behavior of human-beings are collected and assessed for different purposes or goals: to describe a behavior under a certain circumstances, to design a strategy for correcting problem behavior, testing a certain hypothesis about behavior, and evaluating the effectiveness of correcting strategies. Behaviors in Virtual Environment can also be collected and assessed for these goals. Researchers can collect and assess behaviors in Virtual Environments to describe behaviors of human beings in situations that are expensive or time-consuming to be reproduced in the real world, or in situations that potentially harmful for the involved persons. Usability engineers can collect and assess the behavior of a representative user while using a prototype Virtual Reality application in order to identify if there are problems and to suggest correcting strategies to fix identified problems. Trainer can collect and assess the behavior of a trainee in the Driving Simulator Environment in order to evaluate how the trainee performs and suggest correcting feedback if the trainee performs bad.

There are three approaches to collect behaviors that occur in real environment: self report, direct behavioral observation, and psychophysiological measures. The usefulness of these approaches depend on the purpose of the assessment and kind of response to be assessed. Generally three kinds of responses exist that are motoric, physiological, and cognitive. Self report approach collects behavioral data using interviews, questionnaires, and self-monitoring procedures. Even though, it can collect information about the three response modes (overt-motor, physiological-emotional, and verbal-cognitive), it is very effective to assess the verbal-cognitive response system. Direct observation is used primarily in the assessment of the overt-motor response mode by having the client's behavior directly observed by trained recorders or other individuals. It is extremely useful in the specification of a target behavior and its controlling events and in the evaluation of behavioral measures. Psychophysiological measures are used for behavioral assessment in the physiological-emotional response mode and are the only objective means of assessing covert events.

Before evaluating the applicability or usefulness of the three behavioral observation approaches in the case of Virtual Environments, it is wise to discuss the possible response modes to be observed and assessed when we conduct observational studies of behaviors that occur in Virtual Environments. The reason why we observe and assess behaviors in Virtual Environments and the possible domains of application that benefit most when behaviors are recorded and assessed have been discussed in various parts of this report, especially in the description of the research problem and in the concluding section of chapter two. Here, a discussion is given about which kind of response modes is to be collected and the purpose of assessment in each of the potential domain of application of Virtual Reality technology.

Usability evaluation refers to testing the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction. To conduct usability evaluation, quantitative behavioral data, like amount time needed or sequence of activities necessary to perform a certain task, has to be collected to assess effectiveness and efficiency. These behavioral data are collected from overt-motor response mode. Qualitative and subjective behavioral data are also required to assess the satisfaction aspect of usability, which refers to how pleasant and comfortable is to use a system. These kinds of behavioral data occur in the verbal-cognitive response mode and can only be collected subjectively through interviews or questionnaires. Therefore, self-report and direct behavioral observation approaches can be applied to record and assess the usability of new Virtual Environment prototypes.

The second application domain of Virtual Reality technology, which needs observation methods to record and assess behaviors of end-users or other autonomous entities in a Virtual Environment, is the use Virtual Environments to assess the ergonomics of prototypes of new products or services or workplaces. The goal of ergonomics is to put human needs and capabilities at the focus of designing technological systems and make sure that humans and technology work in complete harmony with the equipment and tasks aligned to human characteristics [8]. Usually the ergonomics is evaluated by having a real product or services and observing the patterns of postures or movements of arms,

legs, and any part of human body while using the product or service. However, the Virtual Environment enable to conduct ergonomics evaluation without necessarily having a real manufactured product or service, a virtual prototype of the new product or service can be used to conduct ergonomics evaluation. Ergonomic data like patterns of posture, and movements of hand, leg or any part of human body can be collected by directly observing the representative potential end-user while interacting with the virtual prototype.

The third application domain of Virtual Environments that require recording and assessment of the behavior of end-users or other autonomous is training simulators. VE training simulators can be applied for complex motor learning such as learning to pilot an airplane, to drive a vehicle, air combat training, to perform surgery. Training skills in VE training simulators can be enhanced by assessing how the trainee behaves and giving appropriate feedback to the trainee. To conduct such assessment a structured observation of the behavior of the trainee in the virtual simulator environment is necessary to gather all the relevant observational data and provide appropriate evaluation techniques for assessment purpose. For example, in driving simulator trainer might want to retrieve relevant information like performance, number of errors or faults, number of collisions, handling of navigation situations, speed control, and etc. VE training simulators are effective to train motor and spatial skills and most of the behaviors occur in the overt-motor response mode. As a result, direct behavioral observation is the most preferable to collect behaviors in VE training simulators.

The fourth application domain of Virtual Environment is in conducting analogue assessment. Behavioral researchers set up virtual stimulus events and environmental settings in a Virtual Environment to describe behaviors of customers, behaviors in hazardous environment like nuclear power plants, and treat physiological or psychological disorders (like fear of height or flying). In such situations, all the three kinds of response modes (overt-motor, verbal-cognitive, and physiological-emotional) can occur. Therefore, self-report, direct observational behavior, and psychophysiological approaches can be applied depending on what kind of behavior is to be collected and assessed. For example, psychophysiological approach can be used to collect behaviors in treating physiological disorders.

As it can be noticed from the above discussions, there is no one best approach to collect behaviors in Virtual Environments. All the three behavioral observation approaches: self-report, direct behavioral observation, and psychophysiological are applicable and their effectiveness depend on the kind of domain of application of Virtual Environment for which behavioral observation is to be conducted. Now that the applicability of the three behavioral approaches in Virtual Environment is assessed, another issue can be raised whether they can be applied exactly in the same way as in the case of real environment. The procedure for conducting self-report and psychophysiological measures to collect and assess behaviors is the same whether the behavior occurs in real environment or in virtual environment. However, conducting direct behavioral observation in virtual environment and in real environment can not be the same due to the different characteristics inherited in virtual environment from that of real environment. In virtual

environment, the actors that are interacting are software objects and is difficult to observe and measure all the activities done by them. Again, what we see in the projected screen during a certain Virtual Environment session is according to the perspective (viewpoint) of the End-user of the Virtual Environment; the observer can not see in different angles. All the responses or activities done by the end-users have no the same meaning or purpose to the responses and activities in the virtual world. There fore, it is necessary to have a different way of observing and recording behaviors that occur in virtual environment that takes into consideration the different characteristics of virtual environments as compared to real environments. This does not mean that the steps done in conducting direct behavioral observation in real environment are not necessary in conducting direct behavioral observation in virtual environment. As it is mentioned in earlier, the steps to be included in conducting direct behavioral observation are defining and quantifying behavior, selecting a recording procedure, making sampling decisions regarding how much data to collect and whether (and how) to restrict the situations in which data are collected. These planning steps should be adopted also in conducting direct behavioral observation in virtual environments, as they do not dictate how exactly to capture and record the behavioral data during the observation session.

Software tools like The Observer and CAMERA are designed to support the direct behavioral observation from the first step of design of observational studies and coding schemes, through collection of behavioral data to the analysis and presentation of results. These software tools can be used to support direct behavioral observation in virtual environments even though the way behaviors are captured and recorded has to be resolved yet. In The Observer tool, behaviors are recorded from live observations or video recorded observations by pressing keys or clicking mouse buttons manually by a human observer and can only record automatically the duration of events. When direct behavioral observation with help of The Observer is to be conducted to collect behaviors in virtual environments, the human observer sees only the scene according to the viewpoint of the end-user and cannot see the scene from different angles; and it is difficult to observe and measure all the activities that occur in the virtual environments as they are represented in an abstract way.

5.2 Insights from User Behavior Observation Methods in Interactive Computer Applications

Obviously traditional interactive computer applications and Virtual Environment applications are similar in some aspects, as they both involve interactions between end-users and computer system even though the degree of interaction is more complex in Virtual Environment applications than in interactive web applications. There are also dynamic interactions among autonomous entities in the case of Virtual Environment that is rare in the case of traditional interactive applications. Because of the similarity in some aspects between traditional interactive applications and Virtual Environments, the concepts and techniques discussed in chapter four are assessed in this section if they are applicable in the domain of Virtual Environment.

As it is mentioned in chapter four, end-user's behavior in interactive application is collected and assessed for conducting usability evaluation. Several usability evaluation methods such as Direct Observation, Interview and Focus Group, Questionnaire, Performance Measuring, and Logging User Actions are already described. The question raised now is can we apply the evaluation techniques that are used to record and assess end-user's behavior in interactive applications for identifying usability problems of Virtual Environment applications. Several research studies have been conducted regarding the applicability of the traditional usability methods in interactive computer applications in testing usability of Virtual Environment applications. The results of the previous research studies have identified limitations of existing usability methods for assessing VE systems because of the fact that most VE user interfaces are fundamentally different from traditional graphical user interfaces, with unique I/O devices, perspectives, and physiological interactions. [Kay & Stanney 2002]. It is suggested that novel aspects of human-VE interaction require new evaluation tools [Kay & Stanney 2002]. Toward this end, Gabbard and Hix (1997) have developed a taxonomy of VE usability characteristics that can serve as the foundation for identifying usability criteria that existing evaluation techniques fail to fully characterize [Kay & Stanney 2002].

The limitation of the traditional usability methods for assessing Virtual Environment systems does not mean that all the techniques, which are used for conducting usability test, are not useful for conducting usability evaluation of Virtual Environment application prototypes. The procedures for collecting behavioral information such as Direct Observation, Interview and Focus Group, Questionnaire, Performance Measuring, and Logging User Actions can still be applied without imitating the content of each step. The development of a taxonomy of VE usability characteristics mentioned in the above is meant to guide the design of the content of each step of the usability evaluation methods discussed so far. In the following, a summary of discussion regarding the applicability of traditional usability evaluation techniques in evaluating usability of VE is presented. The discussion is based on the technical report presented in [Bowman,Gabbard,& Hix 01].

One of the most obvious differences between Virtual Environments and traditional interfaces is the physical environment in which the interface is used. In VEs, non traditional input and output devices are used and users may be standing rather than sitting and they may be moving about a large space using whole body movements. Because of the complexities and distinctive characteristics of VEs, a usability study may require multiple evaluators, different evaluator roles and behaviors, or both. These properties give rise to several issues for usability evaluation: [Bowman,Gabbard,& Hix 01]

- Many VE displays do not allow multiple simultaneous viewers (e.g., user and evaluator), so equipment must be set up so that an evaluator can see the same images as the user. If images are viewed on a monitor, then it is difficult to see both the actions of the user and the graphical environment at the same time, meaning multiple evaluators may be necessary to observe and collect data during an evaluation session.

- A common and very effective technique for generating important qualitative data during usability evaluation sessions is the “think aloud” protocol, which is one of the techniques of direct observation. In some VEs, voice recognition is used as an interaction technique, which makes the think aloud protocol much more difficult and perhaps even impossible. Post-session interviews may help to recover some of the information that would have been obtained from the think aloud protocol.
- Another common technique involves recording video of both the user and the interface. Since VE users are often mobile, a single-fixed camera may require a very wide shot, which may not allow precise identification of actions. This could be addressed by using a tracking camera (additional expense and complexity) or a camera operator (additional personnel). Moreover, views of the user and the graphical environment must be synchronized so that cause and effect can be clearly seen on the videotape. Finally, problem of recording video of a stereoscopic graphics image must be overcome.
- Many VEs attempt to produce a sense of presence in the user; that is feeling of actually being in the virtual world rather than the physical one. Direct observation of users of VEs using projected graphics can cause breaks in presence. Direct observation is impossible to conduct in immersive VE like CAVE as the front of a user can not be seen directly.
- Traditional user interfaces typically require only a discrete, single stream of input (e.g., from mouse and keyboard), but many VEs include multi-modal input combining discrete events, gestures, voice, and/or whole body motion. It is much more difficult for an evaluator to process these multiple input streams simultaneously and record an accurate log of the user’s actions. These challenges make multiple evaluators and video even more important.
- Presence is another example of a measure often required in VE evaluations that has no analogue in traditional user interface evaluation. VE evaluations must often take into account subjective reports of perceived presence, perceived fidelity of the virtual world. Questionnaires have been developed that purportedly obtain reliable and consistent measurements of such factors.
- Evaluations based solely on heuristics that is design guidelines performed by usability experts are very difficult in VEs because of a lack of published, verified guidelines for VE user interface design. However, research studies [Joseph & Deborah 97] have been conducted in characterizing usability criteria specifically for VE that serve as a foundation for having design guidelines for VE user interface design. It is not likely that heuristics will appear until VE input and output devices become standardized. Even assuming standardized devices, however, the design space for VE interaction techniques and interfaces is very large, making it difficult to produce effective and general heuristics to use as the basis for evaluation. Therefore, it is difficult to predict the usability of a VE interface without studying real users attempting representative tasks in the VE.
- Because of the complexity and novelty of VEs, the applicability or utility of automated, tool-based evaluation may be greater than it is for more traditional user interfaces. Automated evaluation could reduce the need for several evaluators in a single session.

From the above discussion, it can be said that direct observation method is not suitable in conducting usability evaluation of Virtual Environment applications. In an immersive VE applications it is impossible to watch directly a user while he/she is interacting. Even if it is possible to have direct observation in some VE applications, the task is very difficult as it requires several evaluators in order to observe the multi modal input streams simultaneously and accurately. Even if Video recording of user's actions is possible, there are still challenges in synchronizing the user and the graphical environment views since they cannot be recorded simultaneously by one camera.

Both interview and questionnaire techniques are still valid and useful in conducting usability evaluation of VE applications specially for collecting qualitative and subjective information that are necessary for testing the ease of use and user satisfaction, important aspects of usability evaluation. However, it should be noted that the nature of the questions and factors raised during interviews and in questionnaires must be designed based on the unique usability characteristics of VE applications.

As it is mentioned in chapter four, Logging User Action technique, which collects automatically statistics about the detailed use of a system, have been used for usability evaluation purposes in interactive applications, particularly in web applications. Because of the shortcoming of using direct observation in conducting usability evaluation in VE applications mentioned earlier, automatic logging of user actions is very useful alternative in collecting information during evaluation sessions. The multi modal and complex nature of user actions in VE applications can be recorded simultaneously and accurately using the Logging User Action technique. As it is described in chapter four, tools have been designed and developed for automatically logging the detailed use of a system by users and to detect usability problems by performing automatic evaluation. One of such tools is WebRemUSINE. WebRemUSINE is an environment in order to perform intelligent analysis of Web browser logs using the information contained in the task model of the application. As it is mentioned before, VE applications have unique characteristics that differs them from traditional interactive applications. As a result, no one can expect to use, for example the WebRemUSINE tool to log automatically user actions in VE applications. However, these tools can be useful in designing Logging User Action technique applicable in VE applications. It is known that [Joseph& Deborah 97], no research has been conducted yet that studies automated data collection and evaluation of user's data in VEs.

Chapter Six: Current Developments Toward Observation Methods in VEs

So far, it has been discussed that Virtual Reality technology has a promising application in various domain areas such as engineering, medicine and health care, training, education, entertainment, and scientific visualization. It has been also discussed that recording of how end-users or other autonomous entities behave in Virtual Environment can have a significant impact in enhancing their use. Recording of the behavior of end-users or other autonomous entities during a certain VE usage session can be used in evaluating the usability of VE application prototypes and assessment of the behavior of end-users or other autonomous entities in VE based training simulators, VE used for conducting observational studies for situations that are too expensive or time-consuming to be reproduced in the real world or situations that are potentially harmful, and VE used for virtual prototyping of new products or services. In this chapter, any developments regarding methods for observing behaviors in Virtual Environment are discussed.

As it has been noticed in several research studies that I reviewed so far, there is no an integrated method to observe and assess the behavior of end-user or other autonomous entities in Virtual Environments. However, research works have been conducted that approach some aspects of observation methods in Virtual Environment and all of them deal with usability evaluation methods applicable only for a specific VE application. Here only the summary of three research works are presented that have relevance toward observation methods in Virtual Environments. The first deals with measuring human performance in Virtual Environment that is mainly used for conducting usability evaluation; the second deals with measuring presence in Virtual Environments; and the last one deals with the development of a novel usability evaluation technique for an Immersive Medical Visualization VE application called Crumbs.

6.1 Human Performance Measurement in Virtual Environments

There are several potential advantages of measuring human performance in VE system. The measurement of human performance should be considered during the initial design of VE systems as part of usability evaluation since efficiency is one the criteria for usable systems. The capability to precisely recreate situations or environments across different sessions at different times or places and to hold important parameters constant in VE systems create a favorable condition to use human performance measurement for skill-level certification and personnel selection applications. Measurement of performance in VE based training simulators can also be used to provide meaningful feedback and knowledge of results to individual users. When devising the performance measurement system for any VE application, two levels of measures should be considered. These levels are the primary level that is determined by the specific application that the VE system is designed to address and the secondary level of measures that support the interpretation of the primary measures. Task taxonomies may provide a starting place in the identification

of categories and dimensions of performance measures relevant to a specific VE application. An ANSI (1993) guide presented a ten-category taxonomy of performance measures: time, accuracy, amount achieved or accomplished, frequency of occurrence, behavior categorization by observers, consumption or quantity used, workload, probability, space/distance, and errors. [Kay & Stanney 2002]

There are at least two performance measurement advantages inherent to VE systems. One is the ability to automatically capture aspects of the user's performance, such as operation of the input control devices. The other is the ability to provide an observer with comprehensive perspectives of the user's actions in the virtual environment. The U.S. Army's Simulation Network (SIMNET) system has demonstrated the powerful performance measurement advantages that VE displays can provide for training applications. SIMNET employed a Stealth Vehicle, invisible to training exercise participants that allowed trainers and exercise controllers to view training exercises from any perspective within the virtual environment. In addition, a training mission can be replayed and viewed from any angle. In addition, the user may be interviewed and provided questionnaires to help interpret what occurred and why. [Kay & Stanney 2002]

6.2 Measuring Presence in Virtual Environments

Presence is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience. [Martijn 03]

Although the benefits of presence have been widely discussed, few researchers have attempted to systematically measure this concept and relate it to possible contributing factors. It could be that sense of presence is simply an epiphenomena of good VE design or possibly even a distraction. Without a valid measure of presence, however, this issue cannot be resolved. Determining how to measure presence may provide a means to establish whether presence does indeed enhance VE system interaction and a greater understanding of the factors that drive this phenomenon may result. [Kay & Stanney 2002]

The most commonly used methods to measure presence in current research are based on subjective ratings through questionnaires. An advantage of questionnaire is that not only subjective sensations during the experience in a VE can be measured. Subjects can also asked to describe the VE and their own physiological and behavioral responses. Recently a more or less systematic approach has been taken to establish reliable and validated questionnaires. The most prominent ones are the questionnaire developed by Slater and colleagues, Presence Questionnaire developed by Witmer and Singer, Igroup Presence Questionnaire constructed by Shubert, and ITC Sense Of Presence Inventory (ITC-SOPI) developed by Lessiter. All these presence questionnaires are all originally developed from certain theoretical views on the concept of presence, which determines the scope of the questionnaires as well as the relevant application domain such as

immersive VR or TV. The questionnaire developed by Slater and colleagues is based on several questions, which are all variations on three themes: the subject's sense of "being there", the extent to which the VE becomes more "real or present" than actual reality, and the locality (the extent to which the VE is thought of as a "place" that was visited rather than just a set of images. [Martijn 03]

On the Presence Questionnaire developed by Witmer and Singer, users can rate their experience in the VE according to four factors: control factors (the amount of control the user had on events in the VE), sensory factors (the quality, number, and consistency of displays), distraction factors (the degree of distraction by objects and events in the real world), and realism factors (the degree of realism of the portrayed VE). Igroup Presence Questionnaire, which contains 75 items, is constructed by combining previous published questionnaires and some newly developed questions on technological and context variables. ITC-SOPI, which contains 63 items, is designed to measure presence that should apply across a range of media, for example also to television and cinema. [Martijn 03]

People tend to respond to mediated stimuli as if it were unmediated when they experience a high level of presence. Examining people's reaction to mediated stimuli could provide an objective measure of presence. Objective measures can be divided into behavioral and physiological measures. Several approaches have been suggested to conduct objective measure of presence: measuring reflex responses such as automatically trying to catch a ball or trying to avoid a rapidly approaching object (proposed by Sheridan); and "class A" measures of presence that measure subjects' responses to virtual cues when subjects are also presented with conflicting real cues (proposed by Prothero). [Martijn 03]

Physiological measures include measuring physiological indicators such as muscular tension, cardiovascular responses, and ocular behaviors. It is recommended to utilize a combination of subjective and objective measures of presence to obtain a comprehensive understanding of presence. [Kay & Stanney 2002]

6.3 Usability Evaluation Techniques

It has been explained in chapter five that the traditional usability evaluation techniques are not suitable and have limitation in evaluating usability of VE interfaces. Several researchers have been attempting to extend or adapt the existing traditional usability evaluation techniques to apply to a specific VE interfaces. Generally, the existing traditional usability evaluation techniques that have been adapted can be generally categorized into two: heuristic evaluation and formative evaluation. Heuristic evaluation is performed by usability experts and identifies potential usability problems by comparing the existing user interface to usability principles or guidelines [Joseph, Kent, Kevin, & Deborah 99]. Formative usability evaluation places representative users in task-based scenarios in order to identify usability problems [Joseph, Kent, Kevin, & Deborah 99]. In Heuristic evaluation, gathering of user behavior is not required; in formative evaluation, however, the issue of how to gather and assess behaviors that occurred while a

representative user performs a certain task-based scenario has to be resolved. Since the goal of this paper is to survey methods of behavioral observation in VE, only formative evaluation technique is discussed.

A novel usability evaluation technique have been developed that guides the usability evaluation of Virtual Environments and it has been applied successfully for assessing the usability of an Immersive Medical Visualization VE called Crumbs [Joseph, Kent, Kevin, &Deborah 99]. Here a summary of the usability method is presented with emphasis on how the formative usability evaluation is carried out. Usability evaluation of Crumbs, a tracking tool for biological imaging, is performed by HCI researchers at Virginia Tech and VPST. The formative evaluation of Crumbs is conducted by following the typical, iterative formative path that includes the following steps: develop user task scenarios, representative users perform scenarios with “think and loud” protocol, evaluators collect qualitative and quantitative usability data, evaluators suggest improvements for user interaction design, and evaluators refine user task scenarios. To aid behavioral data collection while the representative users performs the task-based scenarios; an eye tracking system has been integrated into the formative evaluation studies. Eye tracking systems add a new feature and data source to VEs, and greatly enhance the study of cognitive, physical, behavioral, and psychophysical properties in VEs. The CAVE provides complete control over visual stimuli presented to a user; an eye tracker captures and records a user’s actions and responses to such stimuli. A myriad of open research questions exist for which eye scanning data can offer answers, or at least guidelines. Some of these questions, which are critical to usability of VEs are: how does motion in a VE attract attention, how does color affect user responses in a VE, and what do users look at when maneuvering through or manipulating objects within a VE. [Joseph, Kent, Kevin, & Deborah 99]

Several variables are collected and written to a log file during an eye-tracking session in the CAVE. These include a time stamp, the user’s position within the navigable world, the direction of the line of sight vector, and the pupil diameter. The raw data are interpreted to provide researchers more powerful insight: namely, what virtual objects are actually seen by the user. Further post processing of the raw data enables researchers to capture eye angular velocity, the number of objects per second seen by a user, the amount of time a user spends looking at any object, and any paths or patterns a user’s eye follows. The distance that a user’s eyes must travel can also be tracked, allowing evaluators to assess eye-paths during task performance. It is expected that users forced to continuously look back and forth between objects that are far apart will post longer task times, greater error rates, and greater fatigue and eye strain. Computing quantitative information about where the user looks most frequently allows those performing usability assessments to suggest better positioning of objects in the scene, placement of menus, and soon. The pupil’s diameter can be used to measure the intensity of light in the CAVE. CAVE applications generally do not produce excessively bright images, but rapid changes in light intensity can be uncomfortable. By studying the rate and amount of change in the pupil’s diameter, decisions about brightness, contrast, and color can be made to improve the CAVE interface so that it is more comfortable to use. [Joseph, Kent, Kevin, & Deborah 99]

Chapter Seven: Conclusion and Recommendations

7.1 Conclusion

The purpose of this report paper as it is described in chapter one is to conduct survey study on methods of behavioral observation in Virtual Environments and its main goal is to assess the best technique or method for conducting behavioral observation in Virtual Environments. Here, a discussion on the assessment of the alternative behavioral observations is presented based on the insights gained from the previous chapters in order to suggest the best way for conducting behavioral observation in Virtual Environments. Behavioral observation is the collection of behavioral data while subjects engage in various activities under a certain situation and includes the assessment of the collected behavioral data for various purposes: describing behavior, designing or selecting behavioral correcting strategy, and evaluating the selected behavioral correcting strategy. Behavioral observation can be conducted while subjects engage in various activities during a Virtual Environment session to gather behaviors of end-users or other autonomous entities for different purposes mentioned in the previous chapters.

Behavioral data such as end-user's actions can be collected and assessed during formative usability evaluation of newly designed VE application prototypes. In VE-based training simulators like driving simulators, behavioral data such as the way trainees change gear or follow a car or cross an intersection can be gathered and assessed to enhance the use of the driving training simulator by giving feedback to the trainee and results of how the trainee performs during a training session. When VR technology is used to conduct behavioral studies such as assessing behaviors in hazardous environments and treating psychological and physiological disorders (fear of height or flying), it is important to gather how the participant behaves in a certain Virtual Environment session in order to evaluate the effectiveness of treatment strategies.

In the previous chapters, insights gained in defining how users interact with Virtual Environments have been discussed and techniques or methods of behavioral observation from relevant domains: real environment and interactive computer applications have been discussed and their applicability in the domain of Virtual Environments have been assessed. Based on the insight gained from the previous discussions, we can identify the following techniques that can be potentially applied in conducting behavioral observation in Virtual Environments. These are Direct Behavioral Observation, Self-Report (Interview and Questionnaire), Psychophysiological Recordings, Logging User Actions, and Eye Tracking System.

Direct Behavioral Observation is mainly used to observe behaviors that occur in the overt-motor response mode, which can be observed by human senses, by having the subject's behavior directly observed by trained observers. Usually quantitative behavioral data such as frequency of a certain behavior or duration of a certain behavior are collected using Direct Behavioral Observation technique. However, if thinking-aloud techniques is used, some qualitative behavioral data such as emotional feeling such as the state of happiness, sadness or frustration can be collected. Due to some characteristics

inherited to Virtual Environment applications, Direct Behavioral Observation are not suitable, effective and efficient for observing behaviors that occur in Virtual Environment sessions. Most VE applications use audio signals for output and input purposes. Therefore, it is not suitable to use thinking-aloud technique during direct observation in Virtual Environments to collect some qualitative behavioral data. As it is mentioned in chapter five, all responses or activities done by end-users of VE applications have no often the same meaning or purpose to the responses or activities in the Virtual world. The reason for such situations is because of the fact that usually indirect interaction techniques are used to perform some tasks in the Virtual world. As a result, observers have to observe the detailed use of input devices and the activities that occurred inside a Virtual world separately, which requires more than one observer. Observing end-users in immersive VE applications is difficult since an observer cannot stand in front of an end-user. The only possibility is to use tracking camera to observe the detailed input-devices by end-users.

Another problem with using Direct Behavioral Observation in Virtual Environment sessions is when an observer tries to observe the image displayed to an end-user. First of all, most VE applications do not provide two image display one for an observer and the other for an end-user simultaneously. Secondly, the field of view presented to the Observer is according to the view point of the End-user of VE application and this field of view is very restricted and cannot allow to observe the scene, in which activities are happening during VE session, from different angles of perspective. Thirdly, it is not possible to observe and record all the activities that happened very quickly and which are huge amount in a virtual world during a certain VE session because of the limited sensory abilities of human beings.

Self-Report technique, which includes interview and questionnaire, can be applied without any problems in conducting behavioral observation in Virtual Environments. However, they should be used only as a supplementary method to collect qualitative and subjective behavioral data during pre and post session that are important in interpreting quantitative behavioral data collected during VE sessions. During pre-session, Self-Report technique can be used to gather background information such as age, past experience, educational skills, and etc.; during post-session, subjective and qualitative behavioral data such as experiences, preferences, satisfaction, and opinions on performance, ease of use, learnability, and etc., can be collected.

Logging User Actions is the promising technique that can be applied to conduct behavioral observation in Virtual Environments since it is possible and efficient to capture and record automatically events that occur in a virtual world. This technique have been tried successfully in 2D user interface applications such as web applications to collect user actions while users browse a web site. However, tools tried so far can only capture low system events such as system error messages, key presses, mouse clicks, and positions of mouse pointer. So far, there is no a tool (specific or generic) to capture and record events that happened during VE sessions and it poses a challenge to design and develop such tool since VE applications have 3D user interface and most events are high-

level tasks (behaviors) whose specification are dependent on the specific VE application under consideration.

Psychophysiological recordings and Eye Tracking System can be also applied as a supplementary method in some cases where specific kind of behavioral data are required for certain purpose. Psychophysiological recordings can gather physiological responses such as heart-beat, blood pressure, skin conductance, temperature, muscular tension, and etc,. These kind of behavioral data are useful in VE applications designed for treating physiological disorders and in measuring presence in VEs. Eye Tracking System is a real time measurement of pupil position, pupil size, and eye movements and are useful for conducting usability evaluation of VE application prototypes. Again it can be applied in VE-based training simulators, VE used for ergonomic assessment of virtual prototypes, and VE used for conducting behavioral observation studies if the pattern of eye movements is required to be collected for analysis purpose. For example, if one wants to know the lag time between when a trainee driver sees an object and takes an action such as pressing a brake.

7.2 Recommendations

Based on the conclusion given in the previous section, the following can be recommended. As it is explained already, there is a problem in applying Direct Behavioral Observation to observe behaviors that happened during VE sessions. As it has been mentioned in the previous chapters, Direct Behavioral Observation are used predominantly to collect quantitative and qualitative behavioral data in the domains of real environment and interactive computer applications. To collect quantitative and qualitative behavioral data during VE sessions, the Direct Behavioral Observation technique has to be modified or adapted by incorporating other techniques (Logging User Actions, Eye Tracking System, and Tracking VCR) that add extra functionalities to overcome the shortcomings of the Direct Behavioral Observation techniques. The recommended procedure for conducting direct behavioral observation in VE sessions is as follows:

- Describe the purpose for making observation of VE sessions.
- Identify list of objectives to be achieved (questions to be answered during analysis of collected behaviors).
- Defining and Quantifying relevant behaviors chosen based on the above two steps.
- Selecting recording procedures (continuous recording, instantaneous sampling, or one-zero sampling).
- Making sampling decisions (deciding how much behavioral data to capture and log).
- Capturing and Logging those behaviors which are defined and quantified using either Automatic Logging technique or Eye Tracking System, or Tracking VCR.
- Prepare log data and image displays of VE sessions with user defined view angles and link these data source to The Observer software for conducting further behavioral analysis.

References

- [Vince 95] Vince, John: *Virtual Reality Systems*, Wokingham Addison-Wesley 1995
- [Grigore & Philippe 94] Grigore Burdea and Philippe Coiffet: *Virtual Reality Technology*, John Wiley & Sons, Inc. 1994 USA.
- [Kay & Stanney 2002] Kay M. Stanney: *Handbook of Virtual Environments Design, Implementation, and Applications*, Lawrence Erlbaum Associates, Inc 2002 USA.
- [Shamus & David 1] Shamus Smith and David Duke: *Using CSP to Specify Interaction in Virtual Environments*, Department of Computer Science, University of York (Research paper).
- [Shamus & David 2] Shamus Smith and David Duke: *Virtual Environments as Hybrid Systems*, Department of Computer Science, University of York (Research paper)
- [Doug & Larry] Doug A. Bowman and Larry F. Hodges: *User Interface Constraints for Immersive Virtual Environment Applications*, Graphics, Visualization, and Usability Center, College of Computing Georgia Institute of Technology (Research Paper)
- [Anthony & Karen 86] Anthony R. Ciminero and Karen E. Adams: *Handbook of Behavioral Assessment*, John Wiley & Sons, Inc. 1986, USA.
- [CECP 98] Center for Effective Collaboration and Practice: *Conducting A Functional Behavioral Assessment*, Research Paper, May 1998.
- [PBS] Rehabilitation Research and Training Center Positive Behavioral Support: *Methods of Functional Behavioral Assessment*, Research Paper.
- [Noldus 03] Noldus Information Technology: *Technical Specifications of The Observer*, Wageningen, The Netherlands 2003.
- [Geo& Michael&Daniel 00] Geoffreyi, Michael J.Pazzani, & Daniel: *Machine Learning for User Modeling*: Research Report, May 2000.
- [Craig] Craig McQueen: *Usability for Web Applications*, Report Paper, Sage Information Consultants Inc.
- [Ingrid & David 00] Ingrid Zukerman & David W. Albrecht: *Predictive Statistical Model for User Modeling*, Research report, May 2000.
- [Joseph& Deborah 97] Joseph L.Gabbard & Deborah Hix, Ph.D: *A Taxonomy of Usability Characteristics in Virtual Environments*, Department of Computer Science Virginia Polytechnic Institute and State University, November 1997, Research Paper.
- [Bowman,Gabbard,& Hix 01] Bowman, Doug A, Gabbard J. and Hix, D: *Usability Evaluation in Virtual Environment: Classification and*

- [Martijn 03] *Comparison of Methods*, Technical Report TR-01-17, Computer Science, Virginia Tech, 2001.
- [Joseph, Kent, Kevin, &Deborah 99] Martijn Jeroen Schuemie: *Human-Computer and Presence in Virtual Reality Exposure Therapy*, SIKS Dissertation series No. 2003-03
- [Joseph L. Gabbard, Kent Swartz, Kevin Richey, and Deborah Hix 99] *Usability Evaluation Techniques, A Novel Method for Assessing the Usability of an Immersive Medical Visualization VE*, Department of Computer Science, Virginia Tech, 1999.

Electronic Sources:

- [3] <http://giove.cnuce.cnr.it/screenshots.html>
- [4] http://www.ehto.org/ht_projects/vrepar/sensing.htm
- [5] <http://www.behavior.org/behavior/>
- [6] <Http://www.york.ac.uk/inst/ctipsych/web/CTI/WebCiP/Hoogland.html>
- [7] http://www.viterbo.edu/personalpages/faculty/DWillman/p230/p230_nat_obs.htm
- [8] <http://www.ergonomics.org.uk/ergonomics>
- [9] <http://www.org/cdrom/papers/395>
- [10] <http://jthom.best.vwh.net/usability>
- [11] <http://www.ucc.ie/hfrg/questionnaires/sumi>
- [12] <http://www.pages.drexel.edu/~zwz22/FramedLi.htm?Interview.htm>
- [13] <http://www.cs.umd/~zzj/usabilityHome.html>
- [14] <http://www.pages.drexel.edu/~zwz22/FramedLi.htm?Logging.htm>
- [15] <http://www9.org/final-posters/poster20.html>
- [16] <http://www.comp.lancs.ac.uk/computing/users/dixa/papers/hci92-goms/>