The Validity of Presence as a Reliable Human Performance Metric in Immersive Environments

Roy S. Kalawsky

Advanced VR Research Centre, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK, Tel +44 (0)1509 223047, Fax +44 (0)1509 223940, Email: r.s.kalawsky@lboro.ac.uk, WWW: http://sgi-hursk.lboro.ac.uk/~avrrc/index.html

1. Introduction

Advances in interactive media technologies have reached the stage where users need no longer to act as passive observers. As the technology evolves further we will be able to engage in computer based synthetic performances that mimic events in the real or natural world. Performances in this context refers to plays, films, computer based simulations, engineering mock-ups, etc. At a more fundamental level, we are dealing with complex interactions of the human sensory and perceptual systems with a stimulus environment (covering visual, auditory and other components). Authors or rather directors of these new media environments will exploit computer generated perceptual illusions to convince the person (participant) who is interacting with the system that they are present in the environment. Taken to the limit the participant would not be able to distinguish between being in a real or natural environment compared with being in a purely synthetic environment. Unfortunately, the performance of current technology falls considerably short of even getting even close to providing a faithful perceptual illusion of the real world. Eventually, VR authoring tools will become more sophisticated and it will be possible to produce realistic virtual environments that match more closely with the real world.

Some VR practitioners have attempted to collect data that they purport to represent the degree of presence a user experiences. Unfortunately, we do not yet have an agreed definition of presence and seem undecided whether it is a meaningful metric in its own right. This paper will examine the dangers of attempting to measure the 'presence' of a VR system as a one-dimensional parameter. The question of whether presence is a valid measure of a VR system will also be addressed.

2. Presence - The 'Perceptual Sense of Being' in a Virtual Environment

An important differentiating characteristic of VR systems compared with other human-computer interfaces are their ability to create a sense of 'being-in' the computer generated environment. Other forms of media such as film and TV are also known to induce a sense of 'being-in' the environment. Some VR practitioners have tended to use the term presence to describe this effect (Sheridan, 1992), (Heeter, 1992), (Kalawsky, 1993b), (Zelzter, 1994), (Hendrix and Barfield, 1996a). This means that people who are engaged in the virtual environment feel as though they are actually part of the virtual environment.

In order to be perceptually present in a virtual environment it is first important to understand what it means to be present in the real-world. A good example of real world experience is that of a roller coaster ride. The sensory inputs to the rider on the roller coaster will be through vision, sound, smell and proprioception. One will experience air rushing over the face and the sound of the roller coaster and other riders screaming and shouting. The sense of vibrations and extreme inertial forces when accelerating, turning and descending, etc will be very real. It is obvious that most people will also experience intense emotions involving fear and excitement. A video of a roller coaster ride there are only two sensory inputs, vision and sound. Both of these sensory inputs will be much less real than the real world. Stereoscopic depth cues will be absent from visual and auditory information. Temporal lags introduced by the video system will further reduce the visual fidelity of the experience. Some sensory cues may even be contradictory. The body will feel comfortable in a normal seating posture but the visual cues (with reference to the horizon) will be indicating that the body is anything but stable. If the roller coaster is experienced in an IMAX cinema the reaction of others will have an effect. Some people report that they can suppress the sense of presence by weakening or strengthening their awareness. Upon receipt of sensory information, some people can fill in gaps to create a better or enhanced sense of what is happening. For example, people who have previously experienced a roller coaster ride would experience a different state of awareness than someone who had never experienced the ride. This implies that previous experience may affect the sense of presence.

2.1 Intersensory Interactions
Traditionally, sensory modalities have been investigated in isolation from one another. It has been suggested by (Sherrington, 1920) that all parts of the nervous system are connected together and no part is capable of reaction without affecting or being affected by other parts. This means that examination of part of the system will inevitably lead to an incomplete understanding of the perceptual experience. Intersensory interaction relates to the perception of an event when measured in terms of one sensory modality which is changed in some way by the concurrent stimulation of one or more other sensory modalities. Given the nature of human sensory system there is great diversity in the intersensory interactions that can be experienced and this adds to the difficulty in understanding what is happening.

2.1.1 Spatial Location
There are at least four sensory modalities that are capable of providing spatial information to the human being. These are visual, auditory, tactile and proprioception. The visual sensory modality is the most spatially acute of the spatial modalities with a resolution acuity of about 1 min of arc. In contrast, the ability to spatialise a 1kHz tone placed in front of the participant's head at varying angular distances from the median plane of the head gives a minimum angle of about 1°. Tactile acuity is a very difficult thing to define since it depends which part of the body is being stimulated. The tongue has a two-point threshold of about 1mm.

2.1.2 Orientation
There are four sensory modalities that support the perception of orientation: visual, tactile, proprioception and vestibular sense. Prioproception is a very powerful mechanism for conveying a sense of body orientation though it has been shown that with time the body can adapt to unusual positions and this can lead to false orientation cues being perceived. The visual and vestibular senses are extremely accurate in conveying a sense of orientation of gravitational direction. This is one of the reasons why the perceptual system can make serious errors in orientation judgement if one of these two sensory modalities is missing or conflicting with the other. Misperception of the body and the gravitational direction vector can cause a shift in auditory localisation cues (Graybiel and Niven, 1951). This phenomenon is known as the audiogravic illusion.

2.1.3 Egocentric localisation
Egocentric localisation is the ability of the human to perceive the direction and distance of objects relative to the observer. Egocentric localisation is achieved by the visual, auditory, tactile and proprioception sensory modalities. It is usual for several of these modalities to act together to give an accurate sense of localisation.

In the real world it is common for several of the sensory modalities receive simultaneous stimulation in a way that reinforces a common multi-modal perception. However, in the virtual environment system it is possible that one or more of the sensory modalities will receive incorrect stimulation due one of the sensory channels not being provided. This phenomenon is sometimes referred to as intersensory bias.

Whilst we can examine sensory interaction and relate this to specific human capability the term presence has defied all attempts to define it in a quantifiable manner. There is clearly a coupling between the senses and the phenomenon of presence (Gilkey, 1995). Gilkey has examined the level of presence experienced by suddenly deafened adults. Suddenly deaf adults frequently complain of a sense of un-connectedness with their surroundings, which supports the view that auditory cues are important for establishing a sense of presence. It is unfortunate that many people make the mistake of assuming that the most important cue in a virtual environment is the visual modality. Even if we concentrated entirely on the visual channel there would still be sufficient auditory cues around in the real world (including self generated auditory noise such as breathing) to limit the sense of sensory depravation that is reported by suddenly deaf people. Interestingly it has been reported by (Gillingham, 1992) that acoustic isolation and lack of auditory cues may account for spatial disorientation.

The term immersion is also sometimes used erroneously to describe the experience of presence. The term immersion in fact refers to the extent of peripheral display imagery. If the display presents a full 360° information space then we are dealing with a fully immersive system. However, if the extent of the display is less than this then we have a semi-immersive system. The term non-immersive is usually reserved for desk-top VR systems. To avoid confusion it is best to associate immersion with the technology characteristics of the display. Unfortunately, these terms are not interchangeable and refer to quite different things. Presence is essentially a cognitive or perceptual parameter whilst immersion essentially refers to the physical extent of the sensory information and is a function of the enabling technology.

3. Perceptual Conflicts - Phantom Illusions

Gibson (Gibson, 1986) has mentioned the notion of co-perception of one's own movement, in other words awareness of locomotion. The visual system "is kinaesthetic in that it registers movements of the body just as the muscle-joint-skin system and the inner ear". In the real world, the visual system perceives information about the environment and one's own self in that environment. Our whole perceptual system behaves in this manner and processes many reinforcing cues from the environment. It is better to think of these cues as reinforcing since they are all contributory rather than some current views that suggest these cues provide a degree of redundancy. Visual kinesis is a powerful perceptual process as can be evidenced by a wide-angle panoramic projection screen. It is quite easy to produce very convincing and compelling visual cues that give the participant a sense of self-locomotion. The visual experience can appear to untrained people as a very vivid illusion of reality even though the participant is anchored to the floor. A similar illusion can occur whilst sitting on a stationary train in a station and an adjacent train pulls away. Sometimes, you become convinced that you are moving and the other train is stationary. It comes as a surprise when you discover that you are in fact stationary. What is very interesting with these experiments is the way visual cues can override cues from the vestibular system. Although it is tempting to isolate a particular sensory modality when trying to explain perceptual phenomena it is problematical.

Someone who is completely blind would argue that they can 'see' the environment through auditory, haptic and kinaesthetic cues. Indeed, when is deprived of the visual channel you soon become aware how extremely the other modalities are. By allowing, the person to move their heads and move within the environment proprioception fills in much of the information that would normally be provided by the visual channel. The presence of all sensory modalities removes some of the ambiguities that can occur with a reduced set of sensory inputs.

Great care must be taken not to infer that everyone behaves in the same manner. Some people are far more sensitive and can compensate for conflicting sensory cues than others.

In the majority of experiments conducted in presence, the experimenters do not address the issue of sensory conflict. It is quite possible that our real-world experiences which are based on a full set of sensory cues do not readily map onto our sense of presence in a sensory deprived computer generated environment. Not only is our sensory system deprived of certain perceptual cues there also may be sensory conflicts, which arise from issues such as lags or temporal anomalies in our system.

Our experiences or priori knowledge of real world systems can greatly influence our internal representation of a sense of being present in an environment. For example, test pilots are used to dealing with tasks in a fixed based (no motion cue) simulator and transferring the experience to the real world. However, it has been established that most combat pilots perform better in simulated missions compared with real battle situations. Obviously, risks are much easier to take in a simulator than in the real world. As a converse argument combat pilots sometimes make different decisions when under combat stress due to a different level of adrenaline. Unless people are carefully trained (and it is very difficult to determine if this can actually be done) then there is a great danger that the subjective evaluation techniques may not be sensitive to the same parameters for each of the experimental participants.

A computer-generated environment can affect the participants experiences in a very profound way by allowing events or situations to be experienced that cannot be achieved in the real world. For instance, it is easy to transport someone to a different temporal domain where events can be slowed down or speeded up compared to real time. In these situations, it is not practical to try and map this onto a real world experience. Consequently, researchers should be very careful when using terms such as low and high presence.

A crude but repeatable measure for presence would be to count the number of sensory inputs that are missing from the virtual environment compared with the real environment (Sheridan, 1992). Unfortunately, even this approach is flawed because each sensory modality does not contribute equally to the sense of presence. It is also likely that individual contributions will change over a period. For example, it is well known that people can become desensitised to certain stimuli.

4. Real World Versus Virtual Environment

There is considerable merit in being able to compare performance in the real world against performance in a virtual environment, especially if the virtual environment is mimicking the real world in some way. This means that metrics developed for the real world case can be deployed in the virtual environment. However, this presumes human performance is the same in real and virtual environments. This factor is very important for training applications where a virtual environment is used to train a particular skill and the skill has to be transferred into the real world. Skill transfer is a very important factor but equally human behaviour and performance in the virtual environment is very important. Conflicting sensory cues could actually modify the user's performance in the virtual environment in a detrimental or beneficial way. In some cases the ability to present only a subset of real world attributes might actually improve the training process. Pilot training is a good example where basic procedural tasks can be taught without the student having to worry about flying the aircraft at the same time. These training systems are known as part task trainers. To avoid many training transfer issues the part task trainer is made as real and representative as possible.

It is possible to extend this idea by investigating the quality of a virtual environment in terms of the tasks to be undertaken. If we gather data on the user's performance in a real environment and the user's performance in a virtual environment then we have some measure of the quality of the virtual interface.

5. The Quest for Understanding Presence

There has been insufficient research into the causes of presence to be able to discuss them definitely and accurately. "There is no scientific body of data and /or theory delineating the factors that underlie the phenomenon" (Held and Durlach, 1992). Despite this, there is a growing quantity of research that is attempting to derive a single dimension for presence. This research is based on subjective rating techniques. Zeltzer proposed a description of virtual reality in his AIP cube (Zelzter, 1994) which sets out to define the components of a synthetic environment in terms of a coordinate system giving a measure of the quality of the system across three interacting parameters: autonomy, interaction and presence. Zeltzer's cube illustrates the three different axes - defining a coordinate system which can be used as a qualitative measure of virtual environments. Autonomy is defined as the ability of the environment to act and react to simulated events. Interaction is the fidelity with which the environment deals with interactions between its participants both human and synthetic. Presence provides a rough (dimensionless) measure of the number and fidelity of available input and output channels. Zeltzer associated the position of an application within the cube with task performance. He indicated that while clearly, an evaluation of a virtual reality system in these terms is highly task dependent, every design solution for a virtual environment can be characterised within these bounds.

At first sight, Zelter's AIP cube looks to be a good way of describing a particular virtual reality system according to where it fits in the cube. However, it is very difficult to characterise a virtual reality system in this way because of the lack of a clear definition for each axis. In particular, the term presence is very difficult to specify in a simple way that would fit the AIP cube. It is tempting to try to classify attributes of a virtual reality system in this way, and then devise a measurement process but there is a serious danger that the real performance controlling factors of a virtual reality system will not be addressed. Moreover, it is not easy to justify the use of a dimensionless performance parameter if it cannot be measured objectively or subjectively against clearly defined metrics.

To begin to understand where and how to evaluate the user's performance when using a virtual environment it is necessary to look further into the unique properties of the system. Traditional empirical human factors based evaluations such as measuring the display resolution of the system are useful but do not necessarily relate too well in terms of overall user performance. For example, it has been shown that performance in a virtual environment is affected if one of the input modalities is removed (Pausch, Shackleford and Proffitt, 1993). This suggests that if we undertake empirical based evaluations we will not be able to draw too many conclusions regarding an integrated interface. In this context, we need to consider the virtual environment system as an entity and thus treat the system as an integrated interface.

Traditional human factors evaluation techniques do not take into account attributes such as presence and greater interactivity. There have been numerous attempts to produce a single metric representing the degree of presence for a virtual environment and then relate this to some measure of human performance (Slater, Usoh M. and Steed A., 1994). Research which investigated the sense of presence (as yet undefined) within virtual environments as a function of visual display parameters (Hendrix and Barfield, 1996a). The research indicated that people reported higher levels of presence when head tracking was used and stereoscopic visual cues were employed. An increase in field of view also resulted in a reported increase in the level of presence. In reality, these finding are no surprise since we routinely use the cues in the real-world.

There have been many attempts to define a straightforward definition for presence, largely without success. Some VR practitioners try to define different classes of presence such as ego-presence and object-presence (Hendrix and Barfield, 1996a). Indeed, it is tempting to try and derive a simple measure for the amount of presence a particular system is able to provide and then relate this to user task performance. Unfortunately, this approach is flawed because presence is a multi-dimensional parameter that is arguably an umbrella term for many inter-related perceptual and psychological factors. However, it is clear is that presence is a cognitive factor that must be treated differently than other perceptual aspects of a human-computer interface such as brightness or contrast of an image. If presence can correlate usefully with performance and provide the means to achieve effective communication and control in interface design, (Ellis, 1996).

A review of the literature (Airey, Rohlf and Brooks, 1990, Barfield, Zeltzer, Sheridan and Slater, 1995, Ellis, 1996, Fontaine, 1992, Furhammar, 1963, Gibson, 1986, Gilkey, 1995, Gillingham, 1992, Hatada, Sakata and Kusaka, 1980, Heeter, 1992, Held and Durlach, 1992, Hendrix and Barfield, 1996a, Hendrix and Barfield, 1996b, Kalawsky, 1993a, Kalawsky, 1993b, Kalawsky, 1998, Kalawsky, Bee and Nee, 1999, Kuipers, 1982, Regenbrecht, Schubert and Friedmann, 1998, Schloerb, 1995, Sheridan, 1992, Sheridan, 1996, Welch, 1996, Witmer and Singer, 1998), has enabled the following table to be constructed for possible causes (positive effect and negative effect) of presence.

Variable	Contribution
Form Variables – This group includes the more objective parameters	
Sensory outputs	
Number of sensory outputs	Positive (for higher numbers)
Consistency of sensory outputs	Positive when consistent)
Visual outputs – have various dimensions	Strong – see dimensions below
Display size	Positive (for larger proportion)
Viewing distance	Positive (for larger proportion)
Quality of image	Positive (for high quality)
Depth cues	Positive
Camera techniques	Positive
Audible outputs – also has different dimensions	Strong
Other sensory outputs (smells, touch etc)	Can be influential but usually less strong than audio or visual
Body movement and force feedback	Positive when done well
Interactivity of medium	Positive
Visibility/obtrusiveness of medium	Negative
Interference from real world	Negative
Human contact	Positive
Content Variables – Can be both objective and subjective	
Characters and storylines	Positive and negative
Media conventions	Usually negative
Nature of representation	Positive and negative
Media user variables – These are highly subjective and depend directly on the individual	
Willingness to suspend disbelief	Positive
Previous experience	Positive or negative

Table 1: Possible Causes of Presence

6. Issues of Evaluation

A virtual interface is radically different compared to conventional computer interfaces and as such need quite different approaches to performance evaluation. The user's performance is governed by the environment, personal capabilities, individual motivation, the tasks to be performed and the situation under which those tasks are to be carried out. For instance, if the user is performing two tasks simultaneously then performance on a single task might not be the same as if only one task was being undertaken. Whilst it is possible to perform empirical experiments to predict human performance these do not tend to deal with a complex situation where numerous activities have to be undertaken concurrently. An empirical understanding of human performance is important but what is probably more important is an understanding of the overall user performance. This seems strange when a virtual environment system has the potential for so much variability in the design of the interface.

It is easy to overlook that we are dealing with a multi-sensory interface that can provide auditory, kinaesthetic and visual displays. One point to bear in mind during the evaluation process is that the user will experience fewer sensory cues in a VE than in the real world. This inevitably means that our knowledge which relates to the real world may only partially fit the case for virtual environments. Indeed we only need to think of cues such as motion perception to begin to understand the complexity of the problem.

The user of a virtual reality system will generally act inside the environment rather than outside as with other computer based systems. In many ways the flight simulator (one form of virtual environment) has similar attributes. As a result a number of interesting human factors challenges will result. For example, if a user is performing a task in a virtual environment it is quite possible that an experimental evaluator (on the outside) will interfere with the performance of the user. Fortunately, this type of problem can be overcome by careful experimental design. In situations where highly realistic virtual environments are being used, the lack of certain real or redundant sensory cues may have a detrimental effect on the user's performance and subjective experience. An equally important issue is one of perceptual conflict where for example, dominant visual cues may conflict with whole body kinaesthetic cues. This has been known to be the cause of many accidents in the aerospace sector where pilots have tended to believe their own proprioceptive senses rather than aircraft instrumentation. In some instances it will not be possible to avoid such complexities, as the enabling technology will be limited. However, this is where an understanding of empirical human performance becomes important. One way of avoiding the difficulties of human performance evaluation is the development of an evaluation framework. The framework could help formalise the whole process and ensure that a consistent approach is taken.

6.1 Objective Performance Measurement

Objective measures are generally preferred when performing a human factor analysis as they provide a measure of performance against quantifiable parameters. Such parameters include: number of errors made, time to complete task, etc. Objective measures are usually collected on-line and without user intervention. The distinct advantage here is that the user is not subjected to questioning that is intrusive and could interfere with the outcome of the trial. Objective measures of human performance are well established for traditional manual control tasks. Two examples of such tasks, piloting aircraft and driving automobiles, have provided the motivation for much of the work in this area (Wickens, 1986), (Poulton, 1974). There is a clear relationship between the mechanics of flying or driving through a real-world environment and moving the user's viewpoint through a virtual environment. However, the evaluation techniques used by researchers in this area need not be confined to egocentric motion. Indeed, target acquisition and object manipulation may also be evaluated with these techniques (Jagacinski, 1987), (Zhai, Buxton and Milgram, 1994). The application of these techniques to virtual environments still lacks maturity. There are several types of objective evaluation possible in virtual environments, which have originated from the human factors/engineering psychology domain. Let us consider the fundamental control requirements for effective interaction in a virtual environment. Table 2 shows two low level tasks for a virtual interface and the associated user requirements.

	Mode	Requirement
Object Manipulation	Selection	The user must be able to indicate objects of interest both for inspection and for manipulation.
	Translation	The user will tend to require the ability to move objects within the environment.
	Rotation	The user will require the ability to orientate objects within the environment. In many instances the translation and rotation aspects of object manipulation will be coupled.
Egocentric-Motion		The user must be able to move freely through the environment.

Table 2: Low Level Motor Tasks for a VE

Certain applications require more complex high level interactions. For example, object modelling tools enable the user to scale objects interactively. This task may be broken down into the selection and manipulation (typically translation) of control points. Therefore, it may be stated that the actions highlighted above, form the baseline for a generic virtual environment control interface.

Techniques for evaluation of control components have been split between static and dynamic tasks. The static tasks tend to take the form of a docking procedure, where the user is required to match a cursor and target object as quickly and accurately as possible, (Chen, 1988), (Liu, 1992), (Zhai, 1996). Statistical significance can be obtained with this technique by repeating the process a number of times with pseudo-random target. The experimenter would typically record time to complete and accuracy for each trial. The difficulty in implementing this process is in whether to emphasise accuracy or speed of completion when describing the task to the user. Maybe as a result, this technique has occasionally failed to produce significant results when comparing interface components, (Chen, 1988). In the case of Chen et. al. subjective preference proved to be a more reliable technique for comparison.

The dynamic task usually involves the user tracking a target object which is moving in an unpredictable manner (Jagacinski, 1987), (Knight, 1987), (Poulton, 1974), (Wickens, 1986), (Zhai, 1996). The dynamic task is continuously monitored over a period of a few minutes, (Poulton, 1974) suggests between 1 and 4 minutes, per trial). From the data related to cursor and target state (e.g. position and/or orientation) accrued over the duration of the track, a number of different types of analyses may be carried out. Poulton gives a comprehensive review of recommended and not recommended methods of scoring. Much of the work in virtual environments has used RMS error as an over all indication of performance. While this tends to be reasonably reliable when applied over a number of tracks it is a gross measure of performance and tells us nothing of the subtleties of how the interface was being used. A more detailed analysis may be carried out in the frequency domain (Jagacinski, 1987), (Poulton, 1974), (Wickens, 1986). This will show how the user tended to either lead or lag the target and any differences in gain between the target and user's response. A further measure of noise (or remnant) indicates where the user was making movements which were not present in the target's movement. Analysis in the frequency domain is less common in virtual environment evaluation, because the technique has tended to be related to single dimension tracks, where VE interfaces usually deal with multi-dimensional tracks. Also, it is difficult to implement, especially with less accurate responses. Dynamic tasks as a whole tend to provide consistently significant differences in performance across interface components. However, they may not always be applicable. The demand on attentional resources will tend to be higher than with a static task. Clearly where the dynamic task may be suitable for representing dynamic situations, such as in a flight simulator, the static task may be more representative of a CAD system where the attentional demands are lower. It is important at this stage to consider what defines an ideal control interface. (Wickens, 1986) highlights 3 criteria for ideal control, see Table 3.

Control Criteria	
Low Error	The differences between target and cursor are recorded over time. When considering a tracking task the error may arise in a number of different ways. There may be an error in phase and/or gain. Also there is likely to be a certain amount of noise in the system. RMS error analysis is used to establish a gross measure of error.
Stability	This refers to the user's ability to achieve a bounded response to the target movements. The contrary would cause oscillatory behaviour, where the user could not home in on the movement of the target. Stability is measured through linear systems analysis. Typically the experimenter would use a quasi-linear model to represent the subject's response with the controller.
Control Activity	Wickens (1986) states that, "In many circumstances performance is 'better' if the amount of control activity delivered to the system input ... is low".

Table 3: Control criteria for ideal control from (Wickens, 1986)

The key to effective objective performance monitoring for virtual environments is an understanding of the task requirements of the specific application. This will determine the emphasis placed on each of the control components and consequently the relative importance of each of the interface components. From the traditional human-computer interaction field come a number of techniques for breaking down task requirements for simple computer based tasks and analysing human performance at a cognitive and motor level within this framework (this approach is exemplified by Card, Moran & Newell (1983)). As yet, this type of analysis of 2D interfaces does not appear to have been comprehensively extended to dynamic 3D virtual environments. This may be due to the complexities, both cognitive and perceptual, of a task. Instead, researchers have concentrated on evaluating human performance with the interface components (as highlighted above) and using subjective assessment for the completed interface.

6.2 Subjective Performance Measurement

In human factor evaluations it is not always possible or desirable to run experiments where objective performance data is obtained. For instance, in complex experiments it can be very difficult to fully instrument the system under evaluation because of the cost and complexity involved. Instead, subjective measures that reflect subjective opinion have to be collected from trial participants. To the engineer, analysis of subjective data is viewed as a very imprecise process. However, psychological based research methods can be extremely powerful in problematic aspects of human performance. Quite often these methods yield important data that cannot be obtained by any other method. However, to be useful a large number of subjective responses need to be collected in order to achieve statistical significance. As with all psychological based evaluations the sample group must be typical of the user population. Particular care must be taken to ensure that the participant pool is representative of the eventual user group for the VR system. A typical method of obtaining subjective data is via the use of carefully designed questionnaires, attitude scales and surveys. The questionnaire tests for current opinions or patterns of behaviour. Subjective evaluation rating scales have been used for simulation fidelity and workload measurements in virtual environments (Slater, et al., 1994), (Hendrix and Barfield, 1996a), (Kalawsky, et al., 1999). These techniques must be used with great care since the results are not easy to generalise across different raters and different rating situations. A downside of subjective rating scales is the difficulty in developing simple explanatory concepts and stable equivalence classes because of individual differences across the human raters (Ellis, 1996). An important stage in designing subjective evaluation techniques is the issue of reliability and validity.

The evaluator must note that, as subjective measures are, more or less, static (in other words they reflect mean values over a period of time). In complex situations where the task is highly dynamic or where dynamic subjective measures are called for this can be a drawback. It may be possible with some tasks to pause the system and collect subjective data, but in general this method should be avoided because of the risk of task interference. The temptation to encourage the user to provide verbal commentary on their actions should also be discouraged unless it is known that the verbal task will not interfere with the task itself. (Kinsbourne and Cook, 1973) determined that concurrent speaking has an interfering effect on manual tasks. They explained the effect by the fact that verbalisation and right-handed motor activity are controlled from the left cerebral hemisphere, whereas left-handed motor performance is controlled by the right cerebral hemisphere. Similar results were obtained by (Hicks, 1978). Verbal commentary is also likely to affect the sense of presence by drawing the user out of the environment.

Type	**Measures**
Workload	Amount of perceived mental and physical effort required to complete task
Attentional Resources	Amount of attention devoted to particular parts of the task
Task difficulty	Perceived view of the level of difficulty in achieving task
Understanding	Degree of understanding of the requirements of the task

Table 4: Examples of subjective workload

The evaluation of how usable a virtual reality system is represents an important method for identifying and subsequently improving the user interface. Usability is a multi-dimensional term that relates to different aspects of a product or system and involves consideration of factors such as ease of use, flexibility, error handling, provision of help, etc. (Shackel, 1981) stated that: “The usability of a computer is measured by how easily and how effectively the computer can be used by a specific set of users, given particular kinds of support, to carry out a fixed set of tasks, in a defined set of environments”. By controlling the usability engineering process it is possible to achieve the following (Preece, 1994)

- Define usability goals through metrics
- Setting planned levels of usability that need to be achieved
- Analysing the impact of possible design solutions
- Incorporating user-derived feedback in product design
- Iterating through the ‘design-evaluate-design’ loop until the planned levels are achieved

The complexities of the virtual reality system make it difficult to apply traditional tools in the evaluation process and in any event a systematic approach must be carefully followed. Usability evaluation is a very useful technique in the designers toolbox. The inter-dependency of the interface’s attributes mean the evaluation process is particularly challenging. Objective measures sometimes correlate strongly with subjective measures and sometimes not. This dissociation cannot be taken to indicate measurement problems when it comes to subjective performance. In addition, performance and subjective measures are sensitive to different task factors.

7. Developing Content for Virtual Environments

Film and video directors are aware of the techniques that can be used to enhance a sense of presence in the script by exploiting certain visual perception cues. Cartoon animators can exaggerate the effects of motion cues in the near field. By careful use of camera panning techniques it is possible to use phantom cues to create realistic visual effects. Motion parallax is another visual effect that helps dispel the belief. However, in a virtual environment the user is in control of the camera viewpoint and this makes it much more difficult to exploit these effects. With training it is possible to use these visual effects and produce fly-through that really enhance the sense of reality. Conversely, it is possible to employ unnatural movements that totally destroy the sense.

It is conceivable that the manner in which the user navigates through the environment that this action alone is sufficient to affect the sense of presence. This clearly depends on the task and the experience of the user. Content providers will need to understand the importance of being present in one their environments and how best to ensure that the perceptual illusion is maintained. Depending upon the nature of the task it is possible that conflicting cues could reduce or remove the sensation of presence. Obviously, for training applications where training transfer is expected from the virtual to the real world this could be problemmatical.

8. The Dangers of Subjective Measurement based on Poorly Defined Definitions

There are many subjective methods reported in the scientific literature which attempt to employ simple post-test rating scales. These rating scales are based on questions that refer to the extend or degree that the participant feels they are present in the computer generated world. For instance, in (Barfield, et al., 1995) several questions that could be asked that relates to a sense of presence are given. For example, “Please rate your sense of being there in the computer generated world …”. This type of rating test presupposes that each respondent understands what
there”. It is unsafe to assume that everyone will use the same basis for his or her response. This is readily observed in flight simulator systems when novices or naive users are given an opportunity to fly the system. They frequently report that it is just like the real thing. However, when experienced test pilots fly the same system they are often critical of even very small deficiencies in the system that are invisible to the novice user. The flight simulator probably represents the highest fidelity immersive virtual environment is use today by extremely critical users. It is interesting to note that when evaluating the quality of the simulator no-one attempts to measure the degree of presence the pilot gets when using the facility. Although we all have a notion of what we mean by the term presence we are still unable to provide a consistent definition. In the case of the flight simulator, it is usual to examine the task to be undertaken and define a series of metrics that describe the effectiveness of the user. Parameters that relate to the technical performance of the system such as minimum resolvable display element are easily measured and are used in the acceptance of the simulator. Other parameters include measurement of phase lag of a control input to display output (this could be a change in the outside world display or a change in system status. Wherever possible, objective performance measures are taken since these are generally repeatable. Subjective measures are only taken where the parameter being observed cannot be taken by objective measurement. Subjective metrics are particularly important

9. Is There Any Point in Measuring ‘Presence’?

It has been shown that virtual environments need to run as smoothly and fast as possible to increase presence (Slater, et al., 1994) and decrease simulator sickness. This research leads to the conclusion that the designers of VEs need to employ a system resources versus virtual world sensory cue ‘cost benefit analysis’ of some description. Naive HCI decisions may prejudice the holistic performance of the system and lead to negative transfer in training applications.

It is apparent that a general theory of activity is required to investigate a set of low level processes that can be tuned at the interface design stage to increase task performance quantifiably. This is an important step as the assumption is made here that there is, at least on some level, a correlation between activity within a VE and activity in the real-world. With all the task dependent design solutions presented by the literature, there is certainly a generic low-level control structure. Therefore, it is important to contextualise this control structure to appreciate its impact on the design of control interfaces. With the recognition of a bottom level control structure come design criteria, which will indeed be driven by task dependencies. While understanding that subjective opinion is clearly valuable, in a sometimes dynamic and highly interactive environment, the benchmark for performance must include an empirical representation of human activity in a dynamic task. As in the design process, VR may borrow from previous work in other research areas. The application of tried and tested techniques will inevitably yield a formal description of the evaluation process at this level. There does seem to be a great deal of merit in being able to understand the effect of presence in a virtual environment. Even though we may not be able to produce a single robust metric for presence that is universally accepted there is good reason why we should begin to understand the contributory factors that heighten or reduce our sense eof presence. It would be even more useful if we could directly relate user performance to one or more contributory factors of presence. If we had such a metric then we could begin to tailor individual virtual environments to specific applications. In the case where training is performed in the virtual environment we would be able to ensure that the training transfer was effective. It is claimed by (Sheridan, 1992) and (Witmer and Singer, 1998) that the degree of presence is enhanced by being able to interact with a virtual environment. This interaction is not confined to the two or three degrees of freedom that conventional computer interfaces provide. This means that there are likely to be many more factors that affect the sense of presence. Providing we can establish a reliable coupling between presence (or its contributory factors)

10. Rationale behind Situation Awareness and the Measurement of Presence, ρ

Measurement of user performance in a virtual environment typically involves dealing with a range of measures including:- Objective measures (Task demands, Task results and correlated measures - error numbers, achieved task levels etc.), Subjective measures (on-line evaluations, post test evaluation questionnaires, explanation of high stress events), Psycho-physical measures (detection of the stimulus, recognition of the stimulus, stimuli discrimination, magnitude), Physiological measures (heart rate, blood pressure, respiration rate, ECG), Task performance and Learning efficiency. Unfortunately, presence does not have a single physical manifestation that can be measured objectively. This means it is necessary to derive a functional or parametric form for representing presence.

The focus in this paper has been on the impact of cognitive function on presence. In particular, the demand on and supply of attentional resources to achieve a particular level of situation and spatial awareness. A simplified parametric equation for presence, ρ can be summarised as :

$$\rho = (\alpha; \beta; \chi; \beta_{d;}\ C; SA, \Delta SA; \gamma; \mu; \tau, \Delta\tau; \Psi; I_{qty}; I_{qual;}\ T_{e;}\kappa; \sigma; \theta,\phi; d_{m;}\ t,\Delta t)$$

where:

Factors	Function
Demand of attentional resources	α = demand on attentional resource
Supply of attentional resources	β = supply of attentional resource χ = concentration of attention β_d = division of attention C = spare mental capacity
Understanding of situation	SA, ΔSA = Situation awareness and change in situation awareness γ = understanding of situation μ = complexity of situation τ, $\Delta\tau$ = Spatial awareness and change in spatial awareness Ψ= familiarity of situation
Information	I_{qty}= information quantity I_{qual} = information quality T_{e} = elapsed time in environment
Technological factors	κ = sensory modality σ = degree of immersion θ = field of view subtended by the participant's eye ϕ= field of regard of participant d_m = Display mode (binocular, monocular) t = update rate Δt = time lag (propagation delay between event and consequential action)

Table 5: Contributory Factors of Presence

The above parametric equation does not take into account every factor that relates to the subjective feeling of being in a virtual environment. However, by assuming that a virtual environment is required to communicate information to a user, who then is required to perform a series of tasks, it is reasonable to look at the consequential demand on attentional resource and the supply of attentional resource. It is thus possible to make comparisons against task performance. Where the user is required to assimilate the information and understand the situation it is possible to begin to explore the user's situation awareness. This has been proven to be a very powerful indicator of cognitive performance in complex and demanding scenarios. The author has used such techniques in his earlier research on virtual cockpits. However, it is extremely difficult to measure situation awareness because of practical issues of when and how to take subjective measurements. Task related performance parameters can be collected on-line but these do not necessarily relate to cognitive performance or demand. A special technique has been developed which allows collection and rating of situation awareness, SA.

11. Description of the VR Situation Awareness Rating Technique (VRSART)

The mere process of collecting situation awareness data can have an interfering influence on a task if the data is collected during the trial. A technique once favoured required the experimenter to stop the experiment at defined points and run a situation awareness debriefing questionnaire. Unfortunately, this approach modifies overall task performance because the participant's cognitive loading is modified as a result of the interference. Instead, it has been found by experimentation that situation awareness debriefing is more effectively run immediately after the trial. A programme of developing a reliable and robust method of eliciting situation awareness measures is underway at the Advanced VR Research Centre. Phase 2 of VRSART is currently undergoing evaluation and forms part of a more extensive human performance in virtual environments research programme.

11.1 Objectives of the VRSART

The objectives behind the design of the diagnostic tool are:

1. To provide a sensitive computer based diagnostic aid to assist in relating task performance to the contributory factors that lead to a sense of presence,
2. To provide a structured method of determining the impact of presence on the efficiency of a spatially immersive virtual environments,
3. To partition presence factors into specific categories
4. To be a sensitive indicator of problematical areas of the VR user interface,
5. To provide immediate feedback on user performance criteria

11.2 Phase 2 VRSART

VRSART has been extended to address the correlation between demand/supply of attentional resources on situation awareness and the sense of presence. Although task performance is of primary concern the impact of secondary factors (such as situation awareness) on overall user performance is extremely important because it may be possible to tailor the virtual environment in such a way as to enhance situation awareness with a consequential improvement in overall task performance. Phase 2 VRSART deals with six principal factors:- demand on attentional resources, supply of attentional resources, Information, Understanding of situation, Presence and Technology Factors.

Demand
- Demand on attentional resources - what percentage of resource were being demanded
- Unpredictability of situation - to what extent was the situation unpredictable
- Variability of influencing variables - an indication of how many variables were influencing the situation

Supply
- Supply of attentional resources - what percentage of available resources are being supplied
- Degree of readiness - readiness to deal with the situation
- Concentration of attention - an indication of mental effort expended to deal with situation
- Division of attention - percentage of attention time devoted to dealing with situation
- Spare mental capacity - an indication of how much spare mental capacity is left to deal with additional tasks

Information
- Quantity of information - how much content is being received
- Quality of Information - goodness of information being received

Understanding
- Understanding of situation - rating of degree of understanding
- Previous experience - rating of familiarity of situation
- Situation awareness - perception of complete awareness of situation
- Complexity of situation - expression of how complicated the situation was
- Familiarity with situation or previous experience
- I felt isolated and not part of the virtual environment
- I felt disorientated in the virtual environment
- I had a good sense of scale in the virtual environment

Presence
- I got a sense of presence (i.e. being there)

Technology Factors
- The quality of the image reduced my feeling of presence
- I felt a sense of being immersed in the virtual environment
- The display resolution reduced my sense of immersion

A special computerised questionnaire has been developed so that subjective data relating to the above can be collected immediately after a trial and with minimum influence on the participant's subjective ratings. Other data relating to Technology factors (update rate, resolution, field of view etc.) are collected.

11.3 VRSART in the Context of an Evaluation

VRSART has been reliably used as part of the evaluation methodology shown in Figure 1. It is not mandatory to follow this approach when using VRSART but if a quality evaluation is to be conducted then these steps are necessary to capture all the contributing human factors. VRSART must be administered immediately after the trial and before any subjective workload or usability tests have been conducted. Ideally, the participant should complete VRSART before leaving the experimental facility. In any subjective evaluation techniques the matter of delayed rating is very important.

11.4 Known limitations of VRSART

As with any subjective measurement tool, care must be taken with interpretation of the results. Whilst it is possible to employ VRSART by itself it is much better to use it as part of an overall evaluation programme where other user performance data (typically, usability VRUSE, Task workload, Context Analysis, task performance, etc.) is collected. This makes it easier to relate any unexpected results with other quantitative data. Comparison of VRSART data with other data can be problematical if the analysis is performed using a spreadsheet package such as Excel. It is considerably easier to employ a statistical analysis system such as SPSS or Statistica (the author’s choice). Apart from these operational considerations VRSART can be a very useful diagnostic tool in the development of spatially immersive interfaces.

11.5 Verification programme

The validation of any human factors tool is a laborious process necessitating a range of trials and repeat trials until statistical confidence is reached with the technique. VRSART is undergoing development and is being applied to a range of scenarios. The resulting data will assist in the evaluation of the technique’s robustness and evaluation. A formal evaluation methodology and protocol has been developed to ensure that the technique is administered in a systematic and consistent manner. All evaluations are subject to a detailed Context analysis to ensure selection of suitable performance metrics.

11.6 Discussion

Effective development of human-computer interfaces for virtual environments is critically dependent on robust design methodologies and human performance metrics. From a user’s perspective human performance is a significant factor and if poorly implemented can lead to an unacceptable system. By measuring the user’s situation awareness with the interface it is possible to identify shortcomings in operation. Unfortunately, relatively little research has been undertaken on the situation awareness provided by a VR system. Designers and developers of VR systems will find VRSART an extremely powerful diagnostic tool and will be of value to anyone who is producing virtual interfaces because it provides a wealth of information about a user’s viewpoint of the interface.

11.7 Future development of VRSART

VRSART is undergoing further testing and verification in a number of spatially immersive applications. The aim is to eventually provide a fully validated measurement tool.

12. Development of a Perceptual Model for Presence

The increasing number of human factors experiments involving cognitive performance requires urgent development of a method of representing and bringing the data together in a consistent manner. The multi-modal nature of a virtual interface is both strength and a weakness. Its weakness arises from the wide variation in interface technology the system can support and the lack of understanding of the benefits a particular interface’s implementation brings to a specific application. Thus it becomes difficult to specify the best interface for a particular task. Considerable literature exists which deals with the more fundamental aspects of an interface but when more than one of the human senses are used (Hollier and Voelcker, 1997) this literature only partly answers our questions. Clearly, human factor evaluations must be undertaken with respect to the multi-modal nature of the interface. Even if such experiments were conducted, it would still be very difficult to relate these results to other situations because of the complex inter-relation between the data. Due to the complexity of dealing with multi-modal interfaces and the time it takes to perform human factors based experiments there is a requirement for the development of multi-modal perceptual models. Such models would allow rapid iteration around different design solutions in an attempt to produce better systems. Single sensory perceptual models exist but are unable to deal with multi-modal interaction. In a VR system the integration of multi-sensory data is important if a user is to profit from enhanced performance. If a perceptual model as proposed by (Hollier and Voelcker, 1997) existed this would greatly aid our understanding of how and where to optimise any virtual environment. The creation of a reliable perceptual model is an involved process which relies on collecting large quantities of human factors data and developing relationships between the data. Testing and validation of the model requires comparison of model predicted performance against experimental evidence. The quality of the perceptual model gradually evolves over time.

13. Conclusions

The absence of a framework in which to describe the sense of presence in the real world makes it extremely difficult to explain the situation as applied to a computer generated environment. For example, many researchers are content with the term presence being a single dimensional parameter that can be 'measured' by collecting subjective ratings at the conclusion of an immersive experience. It is difficult to see what value can be gained by saying that a group of people feel more present in one environment than another. Indeed, it is problematical to try and derive a set of results from people who have their own definitions of what being present means. A multi-sensory interfaces presents additional complications because of the trade-off between different performance requirements across the different modalities. It is difficult to specify exactly what parameter to trade-off for another. This is where a top down approach might yield some convergence in our understanding rather than an attempt to resolve the problem by treating the interface as presenting quite separate interface modalities. Human factor experiments have shown the impact of factors such as display resolution on task performance. While this information is important from an academic perspective, industrial developers are also interested in higher order human factor issues such as the trade-off of a range of parameters against overall user performance. Key questions are frequently asked such as "What spatial resolution is required at a given update rate to complete a given task without compromising overall task performance". The system developer would then be able to adjust the performance budget of the enabling technology without the risk of engineering an over expensive solution. Having derived the required level of system performance it is possible to look at various human factor trade-offs to fine-tune the system design. The idea of a top down human factor systems view is not new and has been successfully used in the aerospace industry for at least decades.

It is also unsafe to make direct comparison between perceptual experiences in the real-world and virtual environments. The real-world tends to provide extremely rich multi-modal sensory cues that are intimately linked to each other. However, in the case of a virtual environment, technological limitations, system design and lack of knowledge about the critical cues leads to a solution that presents a series of sensory conflicts and in some cases sensory depravation for the user. This can have such a dramatic impact on the participant's perceptual system that missing cues can often be incorrectly perceived as phantom cues.

Unless we can establish a physical or behavioural basis to our definition and measurement of presence we end up with a mathematical entity that has no real use. Key questions to ask in the design of any virtual environment (Ellis, 1996) is whether they can complete the task rather than ask if users feel present in the environment. Can they correctly acquire the necessary information? Do they have the necessary control authority? Can they correctly sequence their sub-tasks? Whether or not measurement of presence will actually provide answers to these fundamental questions has yet to be verified.

The nature of human factor based work means that a considerable amount of research still needs to be done. Above all the rapid pace of technological development is outstripping our current knowledge and we will continually have to refine our models of human performance in virtual environments. The massive range and diversity of all the human factors prevents us from undertaking a complete systematic analysis of all combinations of interface components. At first it seems that to undertake meaningful human factor evaluations involves a very difficult process. It is fair to acknowledge the complexities involved but by taking a top down approach it is possible to yield a substantial amount of useful information. We have to be content with examining the critical issues first, and if time (and funding) permits then deeper investigations should be undertaken. The ultimate quest is for a set of robust evaluation methodologies and metrics that support our required level of understanding.

14. Acknowledgements

This work was supported in part by EPSRC (grant GR/K80693) and the Joint Information Systems Committee (Project JTAP 305) whose support is gratefully appreciated. Support from commercial organisations and in particular BT, Division Inc., MultiGen Inc., SEOS Displays and Silicon Graphics Inc., is gratefully acknowledged.

15. References

Airey, J.M., Rohlf, J.H., and Brooks, F., (1990), Towards image realism with interactive update rates in complex virtual building environments, *Computer Graphics*, vol. 24, pp. 41-50.
Barfield, W., Zeltzer, D., Sheridan, T., and Slater, M., (1995), Presence and performance within virtual environments, in *Virtual*
environments and advanced interface design, I. W. Barfield and T. A. Furness, Ed. New York: Oxford University Press, pp. 473-541.
Chen, M., Mountford, S. J., & Sellen, A., (1988), A Study in Interactive 3-D Rotation Using 2-D Control Devices, Computer Graphics, vol. 22, pp. 121-129.
Ellis, S.R., (1996), Presence of Mind: A Reaction to Thomas Sheridan's "Further Musings on the Psychophysics of Presence", *Presence*, vol. 5, pp. 247-259.
Fontaine, G., (1992), The Experience of a Sense of Presence in Intercultural and International Encounters, *Presence*, vol. 1, pp. 482-490.
Furhammar, L., (1963),*Film Influence on Attitudes*: University of Illinois Press.
Gibson, J.J., (1986),*The Ecological Approach to Visual Perception*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Gilkey, R.H., (1995), The Sense of Presence for the Suddenly Deafened Adult, *Presence*, vol. 4, pp. 357-363.
Gillingham, K.K., (1992), The spatial disorientation problem in the United States Air Force, *Journal of Vestibular Research*, vol. 2, pp. 297-306.
Graybiel, A. and Niven, J.I., (1951), The effect of a change in direction of resultant force on sound localisation: The audiogravic illusion, *Journal of Experimental Psychology*, vol. 42, pp. 227-230.
Hatada, T., Sakata, H., and Kusaka, H., (1980), Psychophysical analysis of the 'sensation of reality' induced by a visual wide-field display, *SMPTE*, vol. 89, pp. 560-569.
Heeter, C., (1992), Being There: The subjective experience of presence, *Presence*, vol. 1, pp. 262-271.
Held, R.M. and Durlach, N.I., (1992), Telepresence, *Presence*, vol. 1.
Hendrix, C. and Barfield, W., (1996a), Presence with Virtual Environments as a Function of Visual Display Parameters, *Presence*, vol. 5, pp. 274-289.
Hendrix, C. and Barfield, W., (1996b), The Sense of Presence within Auditory Virtual Environments, *Presence*, vol. 5, pp. 290-301.
Hicks, R.E., Bradshaw, G.J., Kinsbourne, M. and Feigin, D.S.,, (1978), Vocal-manual tradeoffs in hemispheric sharing of human performance control, *Journal of Motor Behavior*,, vol. 10, pp. 1-6.
Hollier, M.P. and Voelcker, R., (1997), Towards a multi-modal perceptual model, *BT Techol J.*, vol. 14, pp. 162-171.
Jagacinski, R.J., Plamondon, B. D., & Miller, R. A., (1987), Describing Movement Control at Two Levels of Abstraction, Proceedings of Human Factors Psychology;pp:199-247.
Kalawsky, R.S., (1993a), Keynote: A Research agenda for virtual environments, Proceedings of Agard Conference Virtual Interfaces for Aerospace Applications, Lisbon;pp:K1-K7.
Kalawsky, R.S., (1993b), The Science and Engineering of Virtual Reality, Proceedings of Virtual Reality International '93,, London.
Kalawsky, R.S., (1998), New Methodologies And Techniques For Evaluating User Performance In Advanced 3D Virtual Interfaces, Proceedings of The 3D User Interface for the Information Worker, London.
Kalawsky, R.S., Bee, S.T., and Nee, S.P., (1999), Human Factors Evaluation Techniques to Aid Understanding of Virtual Interfaces, *BT Technology Journal*, vol. 17, pp. 128-141.
Kinsbourne, M. and Cook, J., (1973), Generalised and laterialised effects of concurrent verbalisation on a unimanual skill,, *The Quarterly Journal of Experimental Psychology*,, vol. **23**, pp. 341-345.
Knight, J.L., (1987), Manual Control and Tracking, in *Handbook of Human Factors*, G. Salvendy, Ed., pp. 182-218.
Kuipers, B., (1982), The "map in the Head" Metaphor, *Environment and Behaviour*, vol. 14, pp. 202-220.
Liu, A.M., (1992) Depth Cues for Telerobotic Tasks in Virtual Environments, . Berkley: University of California.
Pausch, R., Shackleford, M.A., and Proffitt, D., (1993), A User Study Comparing Head-Mounted and Stationary Displays, Proceedings of IEEE Symposium on Research Frontiers In Virtual Reality.
Poulton, E.C., (1974),*Tracking Skill and Manual Control*. London: Academic.
Preece, J., et al,, (1994),*Human Computer Interaction*: Addison-Wesley.
Regenbrecht, H., Schubert, T., and Friedmann, F., (1998), Measuring the Sense of Presence and its Relations to Fear of Heights in
Virtual Environments, *International Journal of Human-Computer Interaction*, vol. 10, pp. 233-249.
Schloerb, D.W., (1995), A Quantitative Measure of Telepresence, *Presence*, vol. 4, pp. 64-80.
Shackel, B., (1981), The concept of usability, Proceedings of Proceedings of IBM Software and Information Usability Symposium, Poughkeepsie, NY;pp:1-30.

Sheridan, T.B., (1992), Musings on Telepresence and Virtual Presence, *Presence*, vol. 1, pp. 120-25.
Sheridan, T.B., (1996), Further Musings on the Psychophysics of Presence. Presence, *Teleoperators and Virtual Environments*, vol. 5, pp. 241-246.
Sherrington, C.S., (1920),*Integrative action of the nervous system*. New Haven: Yale University Press.
Slater, M., Usoh M., and Steed A., (1994), Depth of presence in virtual environments, *Presence*, vol. 3, pp. 130-144.
Welch, R.B., (1996), The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence, *Presence*, vol. 5, pp. 263-273.
Wickens, C.D., (1986), The Effects of Control Dynamics on Performance., in *Handbook of Perception and Performance*, K. B. A. Kaufman, Ed., pp. 39.1 - 39.60.
Witmer, B.G. and Singer, M.J., (1998), Measuring Presence in Virtual Environments: A Presence Questionnaire, *Presence*, vol. 7, pp. 225-240.
Zelzter, D., (1994), Autonomy, Interaction and Presence, *Presence*, vol. 1, pp. 127-132.
Zhai, S., (1996) Human Performance in Six Degree of Freedom Input Control, : University of Toronto.
Zhai, S., Buxton, W., and Milgram, P., (1994), The "Silk Cursor": investigating transparency for 3D target acquisition, *CHI'94*, pp. 459-464.

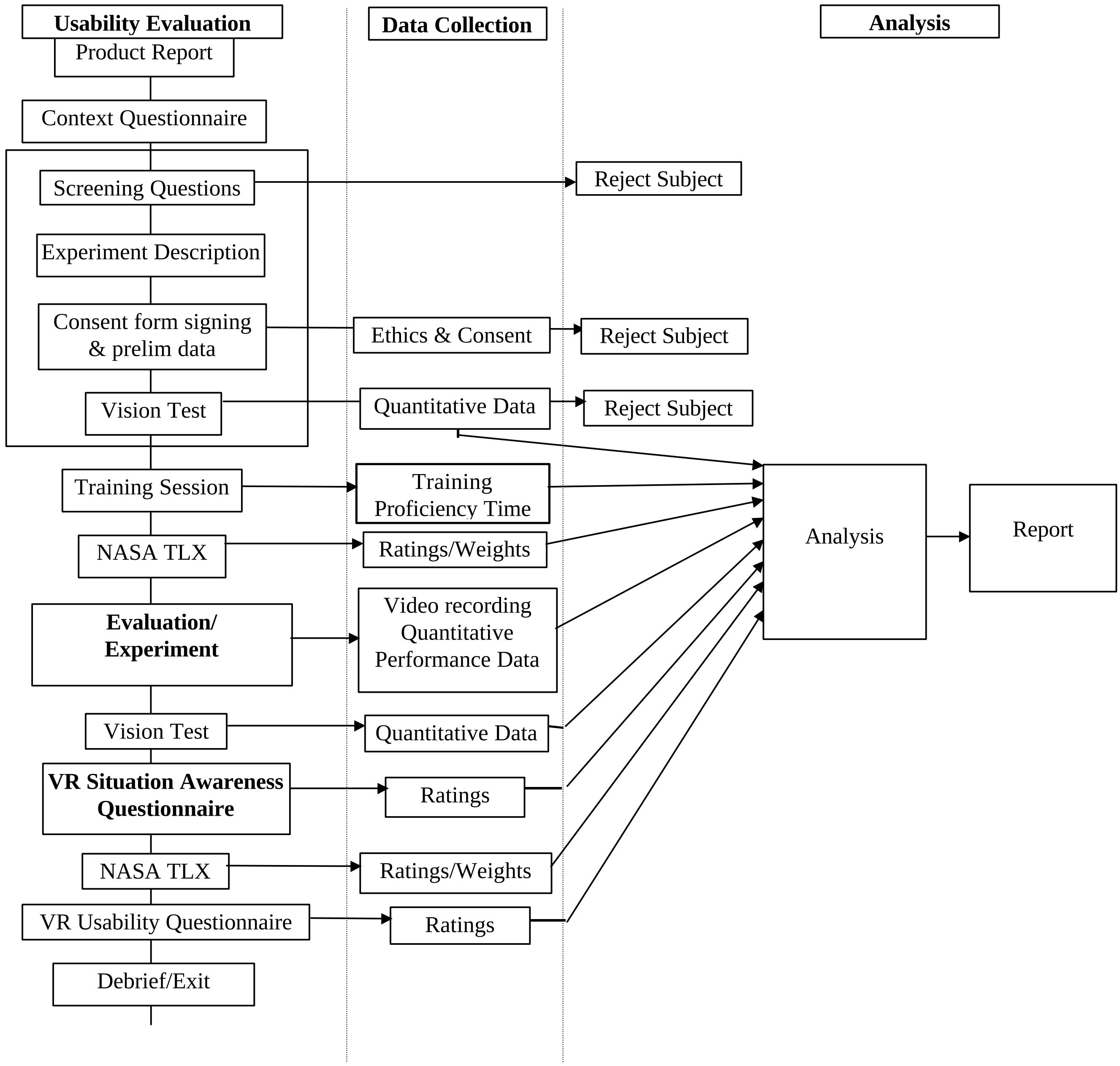

Figure 1: Recommended VRSART Evaluation Framework