7 Measuring Presence: Subjective, Behavioral and Physiological Methods

Brent E. INSKO

Abstract. Virtual reality (VR) systems enable the user to feel as if they are present in a computer generated environment. But how do we determine the extent to which a user feels present in the virtual environment? This chapter examines three categories of methods commonly used for measuring presence; their use in the field, advantages, and disadvantages. Subjective measures rely on self-assessment by the user. Users answer questions such as "How real did the environment seem to you?", "Was the environment like a place you visited, or a series of images presented to you?". Behavioral measures examine actions or manners exhibited by the user that are responses to objects or events in the virtual environment. For example, does the user duck if a virtual object is thrown at his head. Physiological methods attempt to measure presence by gauging changes in the subject's heart rate, skin temperature, skin conductance, breathing rate, etc. In a stress-inducing virtual environment does the user exhibit physiological signs of stress? How do these methods compare when using the criteria of reliability, validity, objectivity, and sensitivity.

Contents

7.1 Definition of presence .................................................. 110
7.2 Subjective measures ...................................................... 110
7.3 Behavioral measures ..................................................... 112
7.4 Physiological measures .................................................. 113
7.5 Comparing presence measures ....................................... 115
7.6 How to choose appropriate measures ............................. 117
7.7 References .................................................................... 118
7.1 Definition of presence

The definition of presence has yet to be agreed upon by researchers. There are two definitions of presence that are most often discussed in the literature. The first is the “sense of being there” in one place or environment (i.e. a virtual environment) even when one is physically situated in another [1]. The second is the “perceptual illusion of nonmediation” [2]. This says that a participant experiences presence, when he fails to perceive or acknowledge that the environment is being presented to him through some type of media (e.g. television, HMDs).

The lack of a single accepted definition, and the subjective nature of the above definitions, leads to difficulty in quantifying a participant’s presence. The first presence measures investigated were post-immersion questionnaires. These asked the user directly questions about presence or factors related to presence. In an effort to create more objective measures, behavioral and physiological methods were introduced. This chapter will examine these three classes of measures, describe several common examples of each, discuss their strengths and weaknesses, and give some guidance on which to use in certain situations.

7.2 Subjective measures

Because of the subjective nature of presence, it is only logical that the first attempts to measure presence relied on the VR participant’s self-reported sense of presence. The most common method of measuring this subjective presence is the post-immersion questionnaire. There have been several different questionnaires developed over the years: Witmer-Singer, SUS, ITC-SOPI, as well as questionnaires specific to experiments, environments, and content. Most questionnaires have participants rate their responses to each question on a numerical scale, though some have examined free response questions [3, 4].

The Witmer and Singer Presence Questionnaire is an attempt at creating a questionnaire valid across media and content [1]. This questionnaire attempts to measure presence by examining factors thought to underlie a person's sense of presence.

These factors are grouped into four categories: Control Factors, Sensory Factors, Distraction Factors, and Realism Factors. The control factors examined included the ability to control the relation of sensors to the environment, the speed at which the system reacts to changes caused by the user, the amount to which a user can anticipate what can happen next in the environment, the "naturalness" of control over the environment, and the ability to modify the physical environment. Sensory factors examine the amounts, coherence, and consistency of information arriving from different senses, the ability to perceive self-movement through the environment and the ability to search the environment actively.

Distraction Factors measure the possible distractions a person may experience in a virtual environment, such as awareness of the real environment, devices used to transmit the virtual environment to the user, and the observer's willingness to focus on the VE stimuli presented. Realism factors measure factors such as environment realism and meaningfulness, as well as disorientation when returning to the real world.

It is important to note that this questionnaire reports the user’s perception of these factors. For example, one question asks, “How much were you able to control events?”.

The problem is that two people may have different responses in the same environment although the system control parameters are the same (Slater [5] argues strongly that this questionnaire measures the user’s perception of system properties, rather than psychological presence).
The Slater-Usoh-Steed (SUS) questionnaire has been developed over a number of years and a number of experiments at the University College of London. The questions are based on variations on one of three themes: sense of being in the VE, the extent to which the VE becomes the dominant reality, and the extent to which the VE is remembered as a place.

Participants rate each of the six questions on a 1-to-7 scale and the number of 6 and 7 responses is counted to produce the score for the SUS. An example question is "During the time of the experience, did you often think to yourself that you were actually in the office space? 1(Not very often) - 7(Very much so)" [6].

The ITC-Sense of Presence Inventory is a questionnaire designed to be relevant across media and across content. It consists of 4 questions rated on a 5-point Likert scale, 1 (strongly disagree) and 5 (strongly agree). These questions relate to 15 content areas deemed relevant to presence on the basis of theoretical and empirical papers. These content areas include: sense of space, involvement, attention, distraction, control and manipulation, realness, naturalness, time, behavioral realism, para-social presence, co-presence, personal relevance, arousal, and negative effects. The questionnaire was tested in a variety of settings: IMAX 2D and 3D movies, a college film night, videos, and playing a racing game on a consumer video game console [7].

Other subjective presence questionnaires include [8-10]. Other subjective measures include the method of paired comparison, where participants are asked to compare their sense of presence in different virtual environments to each other [11] and cross-modality matching where participants can represent their sense of presence by relating it to a sensory modality such as the volume of a radio [12].

7.2.1 Advantages and disadvantages of subjective measures

One advantage of post-immersion questionnaires is their face validity. They appear to measure the concept they propose to measure. Another benefit is the ease of use of questionnaires. Not only are they easy to administer, but they are also easy to grade and interpret. These questionnaires also do not interfere with the user’s experience while in the virtual environment. Questionnaires are also inexpensive to administer and can be setup as online forms for easy compilation.

A major disadvantage of post-immersion questionnaires is the fact that they are post immersion. Not only do these questionnaires not measure the time-varying qualities of presence, but they also may be more influenced by events toward the end of the immersion near the time of questionnaire administration.

During prolonged exposure to a virtual environment, participants may become fatigued or bored thus influencing their responses. To overcome the discrete nature of subjective presence questionnaires, one study examined the use of a continuous subjective reporting device to overcome this disadvantage [13]. The device was a hand-held slider, which the participant could use to indicate his current level of presence and a computer sampled at a constant rate. While this device may not have required a large amount of the participant’s attention, it still intrudes on the participant’s experience, especially if the device were used in a real-walking virtual environment.

Some researchers have identified problems in which the subjective measures produce unstable and inconsistent responses depending on a participant’s prior experience [14].

Many have suggested problems with questions that directly ask participants their sense of presence, because this term is not well known to the general public and requires either the participant to guess or the researcher to explain in detail either of which could influence the participant’s response.
7.3 Behavioral measures

An alternative approach to avoid the subjective nature of questionnaires is to measure participant responses that are produced automatically, without conscious thought, but are sensibly correlated with being present in a particular environment. One such class of measures is behavioral measures. The premise behind these measures is that the more a participant feels present in a virtual environment, the more his responses to stimuli will more nearly match those behaviors he would exhibit in an identical real environment. For example, if a virtual ball was tossed at the user’s head and the user ducks, that would indicate a sense of presence in the virtual environment. Several different measures have been examined.

A study by Freeman et al in 2000 examined postural response as a possible measure of presence. The experiment involved the degree to which participants swayed back and forth while watching a video shot from the hood of a rally car racing around a track. The premise was that the more a participant felt present in the video, the more he would feel as if he were moving and the more postural adjustments he would make. Participants were told to stand as still as possible in front of a screen showing the video either monoscopically or stereoscopically. A magnetic tracking device measured the participants’ movements. There was a positive effect of stereoscopic display on the magnitude of postural movement. Subjective measures of presence were also higher for the stereoscopic presentation. Though subjective presence ratings and postural response were not correlated across subjects, they did corroborate group subjective ratings of presence [15].

Behaviors evoked as responses to stress stimuli have also been examined as potential measures of presence. Studies conducted at the University College of London and the University of North Carolina at Chapel Hill have involved a virtual environment containing a virtual 20-foot precipice (see Figure 7.1) [16-18].

Figure 7.1 Virtual precipice and training room used in experiments at the University of North Carolina at Chapel Hill.
A list of behaviors that are typically exhibited around precipices was constructed. The premise being that the more presence a participant felt, the more real the precipice would seem, and the more pit-avoidance behaviors would be exhibited. Such behaviors included: taking careful baby-steps, leaning away from the pit toward the walls, testing the edge of the ledge with a foot, etc. It was found that these scored behaviors corroborated with the subjective presence ratings between groups [17, 18], though no analysis across subjects has been carried out.

Another type of behavioral measure was proposed by Prothero et al. in which participant responses to virtual cues are measured when participants are also presented with conflicting real cues [19]. Slater et al. [20] conducted such a study in which participants were shown a real radio, then entered the virtual environment with a radio at the same location. During the experiment, the real radio was moved and turned on, and the participant asked to point to the location of the radio. The more present the participant, the more likely he would point to the location of the virtual radio rather than the real radio.

They found a significant correlation between this behavioral measure and a presence questionnaire.

Other type of behaviors that may show participants feel as if they are in the virtual environment include reaching for a virtual object, greeting avatars, turning away or closing ones eyes when presented with an anxiety-provoking scene [21], and startle responses (i.e. flinching and ducking) [22, 23]

7.3.1 Advantages and disadvantages of behavioral measures

Behavioral measures are more shielded from subject bias than subjective measures. Another strength of behavioral measures is their lack of intrusion into the virtual experience.

Behavioral measures can be exposed to experimenter bias. When grading videotaped VR sessions for exhibited behaviors, it is possible for the grader to be aware of the experimental condition and consciously or unconsciously bias the ratings in favor of the results he wants. Another weakness of behavioral measures is the inability to know for a fact that a certain behavior was caused by the experimental condition. A participant losing balance in a VR experience could be from the content of the environment or from the VR system itself. Another problem with behavioral measures is that there is little likelihood that a behavioral measure will be appropriate for all environments and all content. A serious disadvantage of behavioral measures, is that if observed behaviors are scored post-session from videotapes, this will add a considerable amount of time to the completion of the study.

7.4 Physiological measures

There are numerous physiological responses that can be measured. Details of the three most commonly examined and least intrusive measured follow.

Change in heart rate measures the increase or decrease in the number of heartbeats per minute. Heart rate can be measured with an electrocardiogram (ECG) which measures electrical activity across the skin associated with the electrical activity of the heart. Many things can affect a person’s heart rate: stress, fear, exertion, emotion, etc. It increases when a person is under stress and decreases as one relaxes. Heart rate is also affected by the intensity of the emotional response; it increases with position emotions and decreases with negative emotions [24]. In response to unexpected stimuli, heart rate decreases, known as
the orienting effect, and increases with defensive responses, i.e. fight-or-flight response [25].

Change in skin conductance measures the change in the conductivity of a person’s skin. As stress increases, sweat increases on the palms. As sweat increases, the conductivity increases. Higher levels of skin conductance have also been shown during learning [26].

Skin conductance has also been shown to increase with the presentation of an unexpected stimulus[27]. It is measured on the fingertips of the participant. The unit of measure is the Siemen, the inverse of an Ohm, the measure of resistance.

Change in skin temperature measures the change in temperature on the extremities of the body. As stress increases the temperature in the extremities decreases as heat moves to the body’s core. And thus the temperature measured on the fingertip decreases. Skin temperature reactions can be measured by placing a thermistor at the end of one of the fingers and holding it in place with thin porous tape.

More can be found on the causes of these reactions in [28] and [29].

The use of physiological measures in presence experiments has just recently become widespread. Wiederhold et al. compared heart rate and skin conductance to subjective presence ratings. In this study, participants, including phobic and non-phobics, experienced a flying simulation. The results were that the percentage change in heart rate and skin resistance had a high level of correlation with presence [30]. Meehan conducted a series of experiments using the above three physiological measures in the virtual precipice environment as seen in Figure 7.1. His results showed that heart rate was the most sensitive measure of the three, sometimes more sensitive than the subjective measures [17]. More of his results will be presented in Section 7.5.3. Two earlier studies involving these three measures resulted in only skin conductance correlating well with subjective presence measures [21, 31].

Other potentially useful measures might include EEG recordings of cortical activity, electromyogram measurements of facial muscle tensions, respiration rate and blood pressure.

### 7.4.1 Advantages and disadvantages of physiological measures

Physiological measures have several advantages. First, they are more objective than subjective measures and many behavioral measures. They are a continuous measure, so time-varying qualities of presence can be observed. In the Meehan 2001 studies using the monitoring equipment took some training, but added only about five minutes prior to the session for sensor placement. Also most participants reported forgetting about the sensors during the session.

It is important to note that several different stimuli could produce the same changes in physiological measures. It is important in presence experiments with physiological measures that it can be clearly demonstrated what caused the responses. All aspects of the experimental condition and control conditions must be as similar as possible and only differ with respect to the experimental condition. Also important to note is the orienting effect on the exposure to an unexpected stimulus. This can often overwhelm any effects caused by the experimental condition. One method of dealing with this is to expose participants to the stimulus prior to the start of the experiment, thus overcoming the orienting effect.

Physiological levels vary widely from person to person so experiments must measure baseline levels and produce results based on changes compared to that baseline rather than absolute values.

Skin temperature is slow to change. It can take up to five minutes to reach a peak skin temperature. Therefore exposures to the stimulus of several minutes are suggested for
physiological measures using average skin temperature. Because the placement of the skin conductance and skin temperature sensors is on the hand, it prevents that hand from being used in the experiments. In fact, movement of that hand should be limited to prevent data loss from displaced sensors. There is also the cost of the physiological monitoring equipment and a computer to record the data. It is also important to be able to place markers on-the-fly in the recorded data to indicate when certain events occurred in the virtual environment.

7.5 Comparing presence measures

To judge the usefulness and robustness of a measure, the following criteria are used:

- **Reliability** - a measure produces repeatable results, across trials and across participants.
- **Validity** - a measure that measures subjective presence, or at least correlates well with established subjective presence measures.
- **Objectivity** - a measure is minimally effected by and well-shielded from subject bias and experimenter bias.
- **Sensitivity** - a measure can distinguish between various levels of presence.

The following sections detail how each class of measures rate on the above criteria.

7.5.1 Subjective measures

Both the Witmer-Singer [1] and ITC-SOPI [7] questionnaires did their own studies and found themselves to be reliable. The Witmer-Singer and Slater-Usoh-Steed questionnaires have also been used in numerous studies. These subjective measures have face validity. Many questionnaires even explicitly ask "How present were you in the virtual environment?"

Obviously subjective questionnaires are not objective. They are exposed to subject bias, though they are generally not exposed to experimenter bias.

Questionnaires have been the standard measure for presence for many years and have been shown to be sensitive enough to find differences in presence when used to examine: mode of locomotion [16], more sensory cues more presence [32], narrow versus wide FOV [33], etc. However one study did question whether questionnaires could distinguish between a real office and a virtual office [34]. Half of the participants searched for a box in a real office while the other half carried out the same task in a virtual environment that simulated the same office. Immediately following their experience, both the Witmer-Singer and Slater-Usoh-Steed questionnaires were administered. The Witmer-Singer questionnaire failed to show any differences in presence between the two groups, while the Slater-Usoh-Steed questionnaire had a marginally higher mean score for the real compared to the virtual environment.

This brings into question the usefulness of questionnaires when comparing across different and/or no media, such as immersive virtual compared with real, and desktop compared to head-mounted display. However the ITC-SOPI questionnaire was designed to address this cross-media problem, though it is not yet widely used.

7.5.2 Behavioral measures

There has been little research into whether behavioral measures are reliable in virtual environments. Meehan [17] showed that with repeated exposures to a fear-inducing virtual
environment, the number of observed behaviors decreased similar to physiological responses.

Behavioral measures can be valid if it can be shown that the behaviors exhibited by the participants were directly a result of feeling present in the virtual environment. For example, feeling dizzy and losing one's balance can be a behavior exhibited when a person looks over a real precipice. So it stands to reason that loss of balance could be a behavior to watch for in a virtual environment with a virtual precipice. But in a virtual environment, a participant may lose his balance due to other factors such as low frame rate, simulator sickness, etc. Observed fear behavior responses were shown to consistently correlate well with reported behavioral presence, a factor of the UCL presence questionnaire [17]. Insko [18] showed that along with higher changes in heart rate and skin conductivity, the number of observed fear behaviors was higher when participants experienced a virtual precipice augmented with a 1.5 inch wooden ledge as compared with a non-augmented virtual precipice.

Observed behavioral measures can be effected by experimenter bias. Observing from videotapes makes it possible to rewind and make considered judgments. But in some experiments it is impossible for the scorer to be blind to the experimental condition. These measures can also be somewhat exposed to subject bias. In some stress situations, participants may consciously suppress fear behaviors.

In terms of sensitivity, Insko [18] showed that observed behavioral measures in response to a fear-inducing stimulus could differentiate between a virtual precipice, and a virtual precipice augmented with a 1.5 inch physical ledge. Meehan [17] found that observed behavioral measures did not perform well in a study involving three different frame rates.

7.5.3 Physiological measures

The dissertation of Michael Meehan [17] used these criteria to evaluate three physiological measures: heart rate, skin temperature, and galvanic skin response. His studies involved a training room and a "pit" room containing a virtual 20-foot precipice to induce stress in the participants. He showed that the physiological measures reliably distinguished between the training room and the pit room. These measures also reliably decreased over multiple exposures, similar to expected reactions to height in the real world [29, 35].

For validity, change in heart rate was shown to correlate positively with well-established reported measures with significant correlations in one of the three studies. However these physiological measures can be influenced by many factors. So it is important in experiments involving these measures to ensure all aspects of the experiment are identical for each participant except for the experimental condition of interest. It is also important that the exposure to the experimental condition is equal for each participant in the experimental group.

For example, in our experiments at UNC we wanted each participant to walk to the edge of the ledge overlooking the virtual precipice, pause, then walk to the chair on the far side of the room. So our instructions to participants reflected this. Most did not walk to the edge before continuing. So we drew a green dot near the edge and told participants to walk to the dot before continuing with the experiment.

Most participants did not walk out all the way to the dot. Finally we instructed them to walk to a green line drawn on the edge, place their toes over the line and perform a silent count to ten before proceeding. This ensured more equal exposure to the precipice, and showed the importance of piloting experiments.

Physiological measures are more objective than either subjective or behavioral measures; they are generally not consciously affected by the participants, and with limited,
uniform instructions, cannot be affected by the experimenters. Change in heart rate was shown to be a sensitive measure in the studies. As the "goodness" of the virtual environment increases, presence is expected to increase, and thus the change in heart rate should be greater.

The "goodness" of the virtual environment was increased in one study by increasing frame rates, and in another study by augmenting the virtual precipice with a 1.5 inch physical ledge. The change in heart rate was greater in each case as the "goodness" of the virtual environment increased.

7.6 How to choose appropriate measures

So what presence measures should an experimenter choose when conducting experiments in virtual environments? The obvious answer is to use as many as is feasible. If more than one type of measure can be used, it is less likely that any differences in presence between two conditions will be missed. Most of the time, a questionnaire will be appropriate for use in measuring presence in an experiment.

Many early studies into behavioral and physiological measures have examined fear or stress-inducing virtual environments because if measurable results could not be generated in these environments, then there is little hope of results being produced in less intense environments.

The interesting part is how well will these measures extend to less intense virtual environments.

Will the content of any particular environment evoke behaviors and/or physiological responses? In other words, if the content of the virtual environment is an office or other everyday situation, what behavioral or physiological responses can be measured? What kinds of behaviors or physiological responses could be expected from a training environment such as Figure 7.2.

![Figure 7.2 Maze-like virtual environment](image)
If the experimenter seeks to measure differences in methods of presenting the environment independent of content, he should use a stress-inducing environment so that, as stated above, strong behavioral and physiological responses can be expected. It is unclear exactly how behavioral and physiological measures can be used in studies involving environments with different content, for example comparing a virtual office with a virtual pit.

Physiological measures should be useful in a number of different application settings; they should prove useful in clinical applications such as phobia desensitization.

Wiederhold showed that skin conductance was a useful measure when desensitizing phobic with a fear of flying [21]. With a link between skin conductance and learning, skin conductance would be a useful measure when studying learning and training in virtual environments [26]. Physiological measures such as heart rate would be appropriate for many advertising applications to determine which methods of presentation are most appealing. In communication or collaborative environments, behaviors could be used, such as how the participant responds to a virtual human, does the participant smile, offer his hand, etc. As the use of all of the different classes of measures increases, the better we will understand their appropriate use, and hopefully the more we will understand about presence.

7.7 References


