Physiological Measures of Presence in Stressful Virtual Environments

Michael Meehan Brent Insko Mary Whitton Frederick P. Brooks, Jr.
Computer Science Department, University of North Carolina, Chapel Hill
{meehan, insko, whitton, brooks}@cs.unc.edu

Abstract

A common measure of the quality or effectiveness of a virtual environment (VE) is the amount of *presence* it evokes in users. Presence is often defined as the sense of *being there* in a VE. There has been much debate about the best way to measure presence, and presence researchers need, and have sought, a measure that is **reliable, valid, sensitive, and objective**.

We hypothesized that to the degree that a VE seems real, it would evoke physiological responses similar to those evoked by the corresponding real environment, and that greater presence would evoke a greater response. To examine this, we conducted three experiments, the results of which support the use of physiological reaction as a reliable, valid, sensitive, and objective presence measure. The experiments compared participants' physiological reactions to a non-threatening virtual room and their reactions to a stressful virtual height situation. We found that change in heart rate satisfied our requirements for a measure of presence, change in skin conductance did to a lesser extent, and that change in skin temperature did not. Moreover, the results showed that inclusion of a passive haptic element in the VE significantly increased presence and that for presence evoked: 30FPS > 20FPS > 15FPS.

Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - *Virtual Reality*. H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/ Methodology*.

Keywords: Presence, Physiology, Haptics, Frame Rate.

1. Introduction

1.1. Presence and virtual environments

Virtual environments (VEs) are the most sophisticated human-computer interfaces yet developed. The effectiveness of a VE might be defined in terms of enhancement of task performance, effectiveness for training, improvement of data comprehension, etc. A common metric of VE quality is the degree to which the VE creates in the user the subjective illusion of presence – a sense of being in the virtual, as opposed to the real, environment. Since presence is a subjective condition, it has most commonly been measured by self-reporting, either during the VE experience or immediately afterwards by questionnaires. There has been vigorous debate as to how to best measure presence [Barfield *et al.* 1995; Ellis 1996; Freeman *et al.* 1998; IJsselsteijn and de Ridder 1998; Lombard and Ditton 1997; Regenbrecht and Schubert 1997; Schubert *et al.* 1999; Sheridan 1996; Slater 1999; Witmer and Singer 1998].

In order to study a VE's effectiveness in evoking presence, researchers need a well-designed and verified measure of the phenomena. This paper reports our evaluation of three physiological measures – heart rate, skin conductance, and skin temperature – as alternate operational measures of presence in stressful VEs. Since the concept and idea of measuring presence are heavily debated, finding a measure that could find wide acceptance would be ideal. In that hope, we investigated the reliability, validity, sensitivity, and objectivity of each physiological measure.

Figure 1. Side view of the virtual environment. Subjects start in the Training Room and later enter the Pit Room.

1.2. Physiological Reaction as a Surrogate Measure of Presence

As VE system and technology designers, we have sought for a presence measure that is

Reliable – produces repeatable results, both from trial to trial on the same subject and across subjects,

Valid – measures subjective presence, or at least correlates with well-established subjective presence measures,

Sensitive – discriminates among multiple levels of presence, and

Objective – is well shielded from both subject and experimenter bias.

We hypothesize that to the degree that a VE seems real, it will evoke physiological responses similar to those evoked by the corresponding real environment, and that greater presence will evoke a greater response. If so, these responses can serve as objective surrogate measures of subjective presence.

Of the three physiological measures in our studies, Change in Heart Rate performs best. It consistently differentiates among conditions with more sensitivity and more statistical power than the other physiological measures, and more than most of the self-reported measures. It also best correlates with the reported measures.

Figure 2. View of the 20' pit from the wooden ledge.

Change in Skin Temperature is less sensitive, less powerful, and slower responding than Change in Heart Rate, although its response curves are similar. It also correlates with reported measures. Our results and the literature on skin temperature reactions suggest that Change in Skin Temperature would differentiate among conditions better if the exposures to the stimulus were at least 2 minutes [McMurray 1999; Slonim 1974]. Ours averaged 1.5 minutes in each experiment.

Change in Skin Conductance Level yielded significant differentiation in some experiments but was not so consistent as Change in Heart Rate. More investigation is needed to establish whether it can reliably differentiate among multiple levels of presence.

Since Change in Heart Rate best followed the hypotheses, the remainder of this paper will treat chiefly the results for it. For a full account of all measures, please see [Meehan 2001].

1.3. Our Environment and Measures

We use a derivative of the compelling VE reported by Usoh *et al.* [1999]. Figure 1 shows the environment: a Training Room, quite ordinary, and an adjacent Pit Room, with an unguarded hole in the floor leading to a room 20 ft. below. On the upper level the Pit Room is bordered with a 2-foot wide walkway. The 18x32 foot, 2-room virtual space fits entirely within the working space of our lab's wide-area ceiling tracker. Users, equipped with a head-tracked stereoscopic head-mounted display, practice walking about and picking up and placing objects in the Training Room. Then they are told to carry an object into the next room and place it at a designated spot. The door opens, and they walk through it to an unexpected hazard, a virtual drop of 20 ft. if they move off the walkway. Below is a furnished Living Room (Figure 2).

Users report feeling frightened. Some report vertigo. Some will not walk out on the ledge and ask to stop the experiment or demo at the doorway. A few boldly walk out over the hole, as if there were a solid glass floor. For most of us, doing that, if we can, requires conscious mustering of will.

This environment, with its ability to elicit a fear reaction in users, enables investigation of physiological reaction as a measure of presence. If so strong a stress-inducing VE does not produce significant physiological reactions, a less stressful VE won't. This investigation is a first step. Follow-on research should investigate whether less stressful environments also elicit statistically significant physiological reactions.

This remainder section will discuss the physiological measures we tested and the reported measures we used to evaluate validity.

Figure 3. Subject wearing HMD and physiological monitoring equipment in the "Pit Room".

1.3.1. The Physiological Measures

As stated above, we investigated three physiological metrics that measure stress in real environments [Andreassi 1995; Guyton 1986; Weiderhold *et al.* 1998]:

Change in heart rate (Δ Heart Rate). The heart beats faster in stress.

Change in skin conductance (Δ Skin Conductance Level). The skin of the palm sweats more in stress, independently of temperature, so its conductance rises.

Change in skin temperature (Δ Skin Temperature). Circulation slows in the extremities in stress, causing skin temperature to drop.

Each of these measures was constructed to increase when the physiological reaction to the Pit Room was greater.

Δ **Heart Rate** = mean $HR_{\text{Pit Room}}$ – mean $HR_{\text{Training Room.}}$
Δ **Skin Conductance** = mean $SC_{\text{Pit Room}}$ – mean $SC_{\text{Training Room}}$
Δ **Skin Temperature** = mean $ST_{\text{Training Room}}$ – mean $ST_{\text{Pit Room}}$

We first measured heart rate with a convenient finger-mounted blood-pulse plethysmograph, but the noise generated by the sensor moving on the finger made the signal unstable and unusable. We then went to more cumbersome chest-attached three-electrode electrocardiography (ECG). This gave a good signal. Skin conductivity and skin temperature were successfully measured on the fingers. Once connected, users reported forgetting about the physiological sensors – they did not cause breaks in presence during the experiments. Figure 3 shows a subject wearing the physiological monitoring equipment.

1.3.2. The Reported Measures

Reported Presence. We used the University College London (UCL) questionnaire [Slater *et al.* 1995; Usoh *et al.* 1999]. The UCL questionnaire contains seven questions that measure presence (Reported Presence), three questions that measure behavioral presence (Reported Behavioral Presence) – does the user act as if in a similar real environment – and three that measure ease of locomotion (Ease of Locomotion). Responses for each question are on a scale of 1 to 7. Reported Ease of Locomotion was administered for consistency with earlier experiments, but we do not report on it in this paper.

Even though each question is rated on a scale of 1-7, Slater *et al.* use it only to yield a High-Presence/ Low-Presence result. A judgment must be made as to the high-low threshold. Slater *et al.* have investigated the use of 6 and 7 as "high" responses [≥6] and the use of 5, 6, and 7 as "high" responses [≥5] – as well as other constructions: addition of raw scores, and a combination based on principal-components analysis. They have found that [≥6] better followed conditions [Slater *et al.* 1994], and, therefore they chose that construction. We found that the [≥5] construction better follows presence conditions but has lower correlations with our physiological measures. Therefore, in order to best follow the original intention of the measures, irrespective of the lower correlations with our measures, we choose the [≥5] construction. On the study for which data is published, Slater's subjects rarely (<10%) reported "5" values; over 25% of our subjects did. One explanation for this difference in subjects' reporting may be that university students today expect more technically of a VE than they did several years ago and, therefore, are more likely to report lower values (5s) – even for the most presence-inducing VEs.

Reported Behavioral Presence. Three questions asked subjects if they behaved as if present when in the VE. The count of high scores [≥5] on these questions made up the Reported Behavioral Presence measure.

	Multiple Exposures	**Passive Haptics**	**Frame Rate**
Presence in VEs	Does presence decrease with exposures?	Passive Haptics increase presence?	Higher Frame Rate increases presence?
Reliability of Measures	Are repeated measures highly correlated?	Regardless of condition, will the Pit Room evoke similar physiological reactions on every exposure?	
Validity	Do results correlate with reported measures?		
Sensitivity of Measures	Do measures detect any effect?	Do measures distinguish between 2 conditions?	Do measures distinguish among 4 conditions?

Table 1. Questions investigated in each study.

1.4. Methods and Procedures

1.4.1. Experimental procedures.

We conducted three experiments: Effects of Multiple Exposures on Presence (Multiple Exposures), Effects of Passive Haptics on Presence (Passive Haptics), and Effects of Frame Rate on Presence (Frame Rate). Each of the three studies investigated some interesting aspect of VEs and the properties of the physiological measures themselves. Table 1 summarizes all the questions studied. For all studies we excluded subjects who had previously experienced VEs more than three times. The experiments were also limited to subjects who were ambulatory, could use stereopsis for depth perception, had no history of epilepsy or seizure, were not overly prone to motion sickness, were in their usual state of good physical fitness at the time of the experiment, and were comfortable with the equipment.

Multiple Exposures: 10 subjects (average age 24.4; σ = 8.2; 7 female, 3 male) were trained to pick up books and move about in the Training Room – at which time a physiological baseline was taken. Subjects then carried a virtual book from the Training Room and placed it on a virtual chair on the far side of the Pit Room. After that, they returned to the Training Room. The subjects performed this task three times per day on four separate days. We investigated whether the presence-evoking power of a VE declines with multiple exposures. Heart Rate was not successfully measured in this study due to problems with the sensor.

Figure 4. Subject in slippers with toes over 1.5-inch ledge.

Passive Haptics: 52 subjects (average age 21.4; σ = 4.3; 16 female, 36 male) reported on two days. Subjects experienced the VE with the 1.5-inch wooden ledge on one of their two days. The 1.5-inch height was selected so that the edge-probing foot did not normally contact the real laboratory floor where the virtual pit was seen. On their other day, subjects experienced the VE without the ledge. Subjects were counterbalanced as to the order of presentation of the physical ledge. Subjects performed all exposures to the VE wearing only thin sock-like slippers (Figure 4). The task was the same as in the Multiple Exposures study except subjects were instructed to walk to the edge of the wooden platform, place their toes over the edge, and count to ten before they proceeded to the chair on the far side of the room to drop the book. We investigated whether the 1.5-inch wooden ledge increased the presence-evoking power of the VE.

Frame Rate: 33 participants (average age 22.3; σ = 3.6; 8 female, 25 male) entered the VE four times on one day and were presented the same VE with a different frame rate each time. The four frame rates were 10, 15, 20, and 30 frames-per-second (FPS). Subjects were counterbalanced as to the order of presentation of the four

		All exposures				**First Exposure Only (Between Subjects)**			
Study	**Variable**	**Mean**	***P***	**% > 0**	**N**	**Mean**	***P***	**% > 0**	**N**
Multiple Exposures	**ΔSkin Conductance**	2.3 Δ mSiemens	*< .001*	99%	112	**2.9 Δ mSiemens**	*.002*	100%	9
	ΔSkin Temperature	0.6 Δ ºF	*< .001*	77%	94	**1.2 Δ ºF**	*.015*	100%	7
Passive Haptics	**ΔHeart Rate**	6.3 Δ BPM	*< .001*	89%	92	**6.2 Δ BPM**	*< .001*	85%	46
	ΔSkin Conductance	**4.8 Δ mSiemens**	*< .001*	100%	100	4.7 Δ mSiemens	*< .001*	100%	50
	ΔSkin Temperature	1.1 Δ ºF	*< .001*	90%	98	**1.1 Δ ºF**	*< .001*	94%	49
Frame Rate	**ΔHeart Rate**	6. 3 Δ BPM	*< .001*	91%	132	**8.1 Δ BPM**	*< .001*	91%	33
	ΔSkin Conductance	2.0 Δ mSiemens	*< .001*	87%	132	**2.6 Δ mSiemens**	*< .001*	97%	33
	ΔSkin Temperature	0.8 Δ ºF	*< .001*	100%	132	**1.0 Δ ºF**	*< .001*	100%	33

Table 2. Summary of means and significance of differences (Δ) between Training Room and Pit Room. The mean, P-value for the one-sample t-test, percentage of times the measure was > 0, and number of samples are shown. The left side shows the means and significances of all exposures. The right side shows these for only subjects' first exposures. The greater mean is shown in bold.

frame rates. Subjects were trained to pick up and drop blocks in the Training Room and then carried a red block to the Pit Room and dropped it on a red X-target on the floor of the Living Room, a procedural improvement that forced subjects to look down into the pit. They then plucked from the air two other colored blocks floating in the Pit Room and dropped each on the correspondingly-colored Xs on the floor of the Living Room. The X-targets and the green and blue blocks are visible in Figures 1 and 2. In this study, we investigated the effect of several different frame rates on presence and hypothesized that the higher the frame rate, the greater the presence evoked.

In all three studies, the amount of physical activity (walking, manipulating objects) was approximately balanced between the Pit and Training Rooms. This lessened any difference between the two rooms in physiological reaction due to physical activity.

1.4.2. Statistics

In this paper, we define statistical significance at the 5% level, i.e. $P < 0.050$. Findings significant at the 5% level are discussed as "demonstrated" or "shown". To find the best statistical model for each measure, we used Stepwise Selection and Elimination as described by Kleinbaum *et al.* [1998]. As they suggest, to account better (statistically) for variation in the dependent variable (e.g., ΔHeart Rate), we included all variables in the statistical models that were significant at the $P < 0.100$ level.

The analysis of differences in physiological reaction between the Pit Room and the Training Room for all studies (Table 2) was performed with a One-Sample T-Test. The correlations among measures were performed using the Bivariate Pearson Correlation. We analyzed order effects and the effects on presence of passive haptics and frame rate with the Univariate General Linear Model, using the repeated measure technique described in the SAS 6.0 Manual [SAS 1990]. This technique allows one to investigate the effect of the condition while taking into account inter-subject variation, order effects, and the effects of factors that change from exposure to exposure such as loss of balance on the 1.5-inch ledge.

Section 2 details our evaluation of physiological measures as a surrogate for presence. In Section 3, we analyze physiological reactions as between-subject measures. In Section 4, we summarize the results as they pertain to interesting aspects of VEs.

2. Physiological measures of presence

In this section, we discuss the reliability, validity, sensitivity, and objectivity of the physiological measures.

2.1. Reliability

Reliability is "the extent to which the same test applied on different occasions … yields the same result" [Sutherland 1996]. Specifically, we wanted to know whether the virtual environment would consistently evoke similar physiological reactions as the subject entered and remained in the Pit Room on several occasions. Inconsistency could manifest itself as either a systematic increase or decrease in reactions or in uncorrelated measures for repeated exposure to the same VE. In the Multiple Exposures study the condition was the same each time, so this was our purest measure of reliability. We also hypothesized that in the Passive Haptics and Frame Rate studies, regardless of condition, that the Pit Room would also evoke similar physiological reactions on every exposure. We hypothesized that simply being exposed to the Pit Room would cause a greater physiological reaction than the difference between "high" and "low" presence conditions. Therefore, all three studies provide information on reliability.

As we hypothesized, the environment consistently evoked physiological reactions over multiple exposures to the Pit Room. When analyzing the data from all exposures, we found there were significant physiological reactions to the Pit Room: heart rate and skin conductance were significantly higher and skin temperature was significantly lower in the Pit Room in all three studies. Heart rate was higher in the Pit Room for 90% of the exposures to the VE, skin conductance was higher for nearly 95%, and skin temperature was lower for 90%. Table 2 shows the mean difference, t-test, percentage of occurrences where the measure was above zero, and the total count for each physiological measure for each study. It shows results both for all exposures taken together, which is the approach discussed for most of the paper, and for analysis of the first exposure only, which we discuss in Section 4.

We also wanted to know whether the physiological reactions to the environment would diminish over multiple exposures. Since our hypotheses relied on presence in the VE evoking a stress reaction over multiple exposures (2-12 exposures), we wanted to know whether physiological reactions to the VE would drop to zero or become unusably small due to habituation. In fact, ΔSkin Temperature, Reported Presence, Reported Behavioral Presence, and ΔHeart Rate each decreased with multiple exposures in every study (although this effect was not always statistically significant), and Δ Skin Conductance decreased in all but one study. None decreased to zero, though, even after twelve exposures to the VE. Table 3 shows the significant order effects.

A decrease in physiological reaction over multiple exposures would not necessarily weaken validity, since the literature shows that habituation diminishes the stress reactions to real heights and other stressors [Abelson and Curtis, 1989; Andreassi 1995]. Since, however, the reported presence measures, not just the physiological stress measures, decrease over multiple exposures, the decreases may not be due to habituation to the stressor; there may also be, as Heeter hypothesized, a decrease in a VE's ability to evoke presence as novelty wears off [Heeter 1992].

Orienting Effect. In general, each measure decreased after the first exposure. Moreover, for each measure except ΔHeart Rate, there was a significant decrease after the first exposure in at least one of the studies (see Table 3). For physiological responses, this is called an *orienting effect* – a higher physiological reaction when one sees something novel [Andreassi 1995]. Though this term traditionally refers to physiological reactions, we will also use the term for the initial spike in the reported measures.

We attempted, with only partial success, to overcome the orienting effect by exposing subjects to the environment once as part of their orientation to the experimental setup and prior to the data-gathering portion of the experiment. In the Passive Haptics and Frame Rate studies, subjects entered the VE for approximately two

Order Effects	ΔHeart Rate (ΔBPM)	ΔSkin Conductance (ΔmSiemens)	ΔSkin Temperature (Δ°F)	Reported Presence (Count "high")	Reported Behavioral Presence (Count "high")
Multiple Exposures	NA	-0.7 (1st)	**-0.9 (1st)**	-	**-0.7 (1st)**
Passive Haptics	-	-	-	**-0.8 (1st)**	**-0.4 (1st)**
Frame Rate	**-1.0 (Task)**	**-0.8 (1st)**	**-0.3 (1st)**	-	**-0.2 (Task)**

Table 3. Significant order effects for each measure in each study. "(1st)" indicates a decrease after the first exposure only. "(Task)" indicates a decrease over tasks on the same day. There was an order effect for each measure in at least one study. NA is "Not available". Significant results are listed at the P < 0.050 level (bold) and P < 0.100 (normal text). Full details given in [Meehan 2001].

minutes and were shown both virtual rooms before the experiment started. These pre-exposures reduced but did not eliminate the orienting effects.

2.2. Validity

Validity is "the extent to which a test or experiment genuinely measures what it purports to measure" [Sutherland 1996]. The concept of presence has been operationalized in questionnaires so the validity of the physiological measures can be established by investigating how well the physiological reactions correlate with one or more of the questionnaire-based measures of presence. We investigated their correlations with two such measures: Reported Presence and Reported Behavioral Presence.

Reported Presence. Of the physiological measures, ΔHeart Rate correlated best with the Reported Presence. There was a significant correlation in the Frame Rate study (corr. = 0.265, $P < 0.005$) and no correlation (corr. = 0.034, $P = 0.743$) in the Passive Haptics study. In the Multiple Exposures study, where ΔHeart Rate was not available, ΔSkin Conductance had the highest correlation with Reported Presence (corr. = 0.245, $P < 0.010$).

Reported Behavioral Presence. ΔHeart Rate had the highest correlation, and a significant one, with Reported Behavioral Presence in the Frame Rate study (corr. = 0.192, $P < 0.050$), and there was no correlation between the two (corr. = 0.004, $P = 0.972$) in the Passive Haptics study. In the Multiple Exposures study, where ΔHeart Rate was not measured, ΔSkin Conductance had the highest correlation with reported behavioral presence (corr. = 0.290, $P < 0.005$).

The correlations of the physiological measures with the reported measures give some support to their validities. The validity of ΔHeart Rate appears to be better established by its correlation with the well-established reported measures. There was also some support for the validity of ΔSkin Conductance from its correlation with reported measures.

Following hypothesized relationships. According to Singleton, the validation process includes "examining the theory underlying the concept being measured," and "the more evidence that supports the hypothesized relationships [between the measure and the underlying concept], the greater one's confidence that a particular operational definition is a valid measure of the concept" [Singleton *et al.* 1993]. We hypothesized that presence should increase with frame rate and with the inclusion of the 1.5-inch wooden ledge, since each of these conditions provides increased sensory stimulation fidelity. As presented in the next section, our physiological measures did increase with frame rate and with inclusion of the 1.5-inch wooden ledge. This helps validate the physiological reactions as measures of presence.

2.3. Sensitivity and multi-level sensitivity

Sensitivity is "the likelihood that an effect, if present, will be detected" [Lipsey 1998]. The fact that the physiological measures reliably distinguished between subjects reaction in the Pit Room versus the Training Room in every study assured us of at least a minimal sensitivity. For example, heart rate increased an average across all conditions of 6.3 beats / minute (BPM) in the Pit Room ($P < 0.001$) compared to the Training Room in both the Passive Haptics and Frame Rate studies. See Table 2 for a full account of sensitivity of physiological measures to the difference between the two rooms.

Acrophobic patients', when climbing to the second story of a fire escape (with a handrail), waiting one minute, and looking down, averaged an increase in heart rate of 13.4 BPM [Emmelkamp and Felten 1985]. Our subjects were non-phobic, and our height was virtual; so, we would expect, and did find, our subjects' heart rate reactions to be lower but in the same direction.

Multi-level sensitivity. For guiding VE technological development and for better understanding of the psychological phenomena of VEs, we need a measure that reliably yields a higher value as a VE is improved along some "goodness" dimension, i.e., is *sensitive* to *multiple* condition values. We distinguish this from sensitivity as described above and call this *multi-level sensitivity*. The Passive Haptics study provided us some evidence of the measures' ability to discriminate between two "high presence" situations. We have informally observed that walking into the Pit Room causes a strong reaction in users, and this reaction seems greater in magnitude than the differences in reaction to the Pit Room between any two experimental conditions (e.g., with and without the 1.5-inch wooden ledge). Therefore, we expected the differences in reaction among the conditions to be less than the differences between the two rooms. For example, in Passive Haptics, we expected there to be a significant difference in the physiological measures between the two conditions (with and without the 1.5-inch wooden ledge), but expected it to be less than the difference between the Training Room and Pit Room in the "lower" presence condition (without the 1.5-inch wooden ledge). For ΔHeart Rate, we did find a significant difference between the two conditions of 2.7 BPM ($P < 0.050$), and it was less than the inter-room difference for the without-ledge condition: 4.9 BPM. See Figure 5. Figure 6 shows that the differences among the conditions in the FR study are smaller in magnitude as compared to the differences between the two rooms.

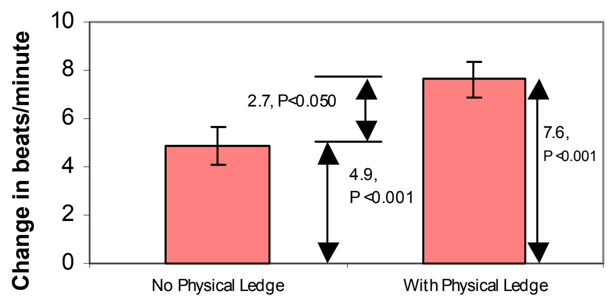

Figure 5. Δ Heart Rate in Passive Haptics study.

In the Passive Haptics study, we investigated the multi-level sensitivity of the measures by testing whether presence was significantly higher with the 1.5-inch wooden ledge. Presence as measured by each of ΔHeart Rate (2.7 BPM; $P < 0.050$), ΔSkin Conductance (0.8 mSiemens; $P < 0.050$), and Reported Behavioral Presence (0.5 more "high" responses; $P < 0.005$) was significantly higher with the wooden ledge. Reported Presence had a strong trend in the same direction (0.5 more "high" responses; $P = 0.060$).

In the Frame Rate study, we investigated the multi-level sensitivity of the measures by testing whether presence increased significantly as graphic frame update rates increased. We hypothesized that physiological reactions would increase monotonically with frame rates of 10, 15, 20, and 30 FPS. They did not do exactly that (see Figure 6). During the 10 FPS condition, there was an anomalous reaction for all of the physiological measures and for Reported Behavioral Presence. That is, at 10 FPS, subjects had higher physiological reaction and reported more behavioral presence. We believe that this reaction at 10 FPS was due to discomfort, added lag, and reduced temporal fidelity while in the ostensibly dangerous situation of walking next to a 20-foot pit [Meehan 2001].

We also observed that subjects often lost their balance while trying to inch to the edge of the wooden platform at this low frame rate; their heart rate jumped an average of 3.5 BPM each time they lost their balance (P < 0.050). Statistically controlling for these Loss of Balance incidents improved the significance of the statistical model for ΔHeart Rate and brought the patterns of responses closer to the hypothesized monotonic increase in presence with frame rate – but did not completely account for the increased physiological reaction at 10 FPS. Loss of Balance was not significant in any other model.

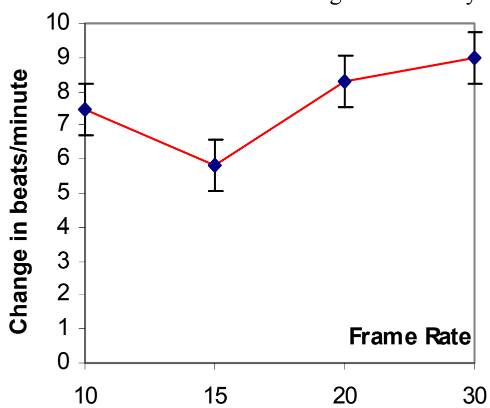

Figure 6. Δ heart rate, after correcting for Loss of Balance, at 10, 15, 20, and 30 frames per second.

Beyond 10 FPS, ΔHeart Rate followed the hypothesis. After we statistically controlled for Loss of Balance, ΔHeart Rate significantly increased between 15 FPS and 30 FPS (3.2 BPM; P < 0.005) and between 15 FPS and 20 FPS (2.4 BPM; P < 0.050). There was also a non-significant increase between 20 FPS and 30 FPS (0.7 BPM; P = 0.483) and a non-significant decrease between 10 FPS and 15 FPS (1.6 BPM; P = 0.134). Reported Presence, and Reported Behavioral Presence also increased with frame rate from 15-20-30 FPS, but with less distinguishing power.

These findings support the multi-level sensitivity of ΔHeart Rate.

2.4. Objectivity

The measure properties of reliability, validity, and multi-level sensitivity are established quantitatively. Objectivity can only be argued logically. We argue that physiological measures are inherently better shielded from both subject bias and experimenter bias than are either reported measures or measures based on behavior observations. Reported measures are liable to subject bias – the subject reporting what he believes the experimenter wants. Post-experiment questionnaires are also vulnerable to inaccurate recollection and to modification of impressions garnered early in a run by impressions from later. Having subjects report during the session, whether by voice report or by hand-held instrument, intrudes on the very presence illusion one is trying to measure. Behavioral measures, while not intrusive, are subject to bias on the part of the experimenters who score the behaviors.

Physiological measures, on the other hand, are much harder for subjects to affect, especially with no biofeedback. These measures are not liable to experimenter bias, if instructions given to the participants are properly limited and uniform. We read instructions from a script in the Multiple Exposures study. We improved our procedure in the later Passive Haptics and Frame Rate studies by playing instructions from a compact disk player located in the real laboratory and represented by a virtual radio in the VE.

2.5. Summary and discussion

The data presented here show that physiological reactions can be used as reliable, valid, multi-level sensitive, and objective measures of presence in stressful VEs. Of the physiological measures, ΔHeart Rate performed the best. There was also some support for ΔSkin Conductance.

ΔHeart Rate significantly differentiated between the Training Room and the Pit Room, and although this reaction faded over multiple exposures, it never decreased to zero. It correlated with the well-established reported measure, the UCL questionnaire. It distinguished between the presence and absence of passive haptics and among frame rates at and above 15 FPS. As we argued above, it is objective. In total, it satisfies all of the requirements for a reliable, valid, multi-level sensitive, and objective measure of presence in a stressful VE.

ΔSkin Conductance has some, but not all, of the properties we desire in a measure of presence. In particular, it did not differentiate among frame rates. We do not have a theory as to why.

Although, ΔHeart Rate satisfied the requirements for a presence measure for our VE, which evokes a strong reaction, it may not for less stressful VEs. To determine whether physiological reaction can more generally measure presence, a wider range of VEs must be tested, including less stressful, non-stressful, and relaxing environments. Investigation is currently under way to look at physiological reaction in relaxing 3D Television environments [Dillon *et al.* 2001].

The height reaction elicited by our VE could be due to vertigo, fear, or other innate or learned response. The reactions are well known in the literature and manifest as increased heart rate and skin conductance and decreased skin temperature [Andreassi 1995; Guyton 1986]. We hypothesized that the more present a user feels in our stressful environment, the more physiological reaction the user will exhibit. What causes this higher presence and higher physiological reaction? Is it due to a more realistic flow of visual information? Is it due to more coherence between the visual and haptic information? Is it due to the improved visual realism? All of these are likely to improve presence. We cannot, however, answer these questions definitively. We can say, though, that we have empirically shown that physiological reaction and reported presence are both higher when we present a "higher presence" VE. Whatever it is that causes the higher reported presence and physiological reaction, it causes more as we improve the VE.

An additional desirable aspect of a measure is ease of use in the experimental setting. We did not record the time needed for each measure, but after running many subjects we can say with some confidence that use of the physiological monitoring and of the presence questionnaire each added approximately the same amount of time to the experiment. It took about five minutes per exposure to put on and take off the physiological sensors. It took about an extra minute at the beginning and end of each set of exposures to put on and take off the ECG sensor – it was left on between exposures on the same day. It took subjects about five minutes to fill out the UCL Presence Questionnaire. It took some training for experimenters to learn the proper placement of the physiological equipment on the hands and chest of the subject – thirty minutes would probably be sufficient.

Another aspect of ease of use is the amount of difficulty participants have with the measure and to what extent the measure, if concurrent with an experimental task, interferes with the task. No subjects reported difficulties with the questionnaires. Only about one in ten subjects reported noticing the physiological monitoring equipment on the hands during the VE exposures. Our experiment, though, was designed to use only the right hand, keeping the sensor-laden left hand free from necessary activity. No subjects reported noticing the ECG sensor once it was attached to the chest. In fact, many subjects reported forgetting about the ECG electrodes when prompted to take them off at the end of the day. There are groups investigating less cumbersome equipment, which would probably improve ease of use, including a physiological monitoring system that subjects wear like a shirt [Cowings *et al.* 2001]. Overall, questionnaires and physiological monitoring were both easy to use and non-intrusive.

3. Physiological reactions as between-subjects measures

We conducted all of the studies as within-subjects to avoid the variance due to natural human differences. That is, each subject experienced all of the conditions for the study in which she participated. This allowed us to look at relative differences in subject reaction among conditions and to overcome the differences among subjects in reporting and physiological reaction.

The UCL questionnaire has been used successfully between-subjects [Usoh *et al.* 1999]. We suspected, however, that physiological reaction would not perform as well if taken between-subjects. We expected the variance among subjects would mask, at least in part, the differences in physiological reaction evoked by the different conditions. We investigated this hypotheses by analyzing the data using *only the first task* for each subject – eliminating order effects and treating the reduced data sets as between-subjects experiments. That is, we treat each experiment as if only the first task for each subject was run. This means that the analysis uses only 10 data points (10 subjects – first exposure only) for the Multiple Exposures study, 52 data points for the Passive Haptics study, and 33 data points for the Frame Rate study.

Reliability between-subjects: Physiological reaction in the Pit Room. Even between subjects, we expected that there would be a consistent physiological reaction to the Pit Room, since we expected such a reaction for every exposure to the VE. We expected the significance to be lower, however, because of the reduced size of the data set. We found exactly that. The right half of Table 2 shows the values of the physiological measures averaged across conditions for the between-subjects analysis. As compared to the full data set, the between-subjects data have lower significance values, but subjects still have strong physiological reactions to the Pit Room. Table 2 demonstrates that the physiological orienting effects caused the averages for the first exposures to be *higher* than for the full data set.

Validity between-subjects: Correlation with established measures. We expected correlations with the reported measures to be lower when taken between subjects since there were fewer data points and individual differences in physiological reaction and reporting would confound the correlations. This was the case. No physiological measure correlated significantly with any reported measure when analyzing between-subjects.

Multi-level sensitivity between-subjects: Differentiating among presence conditions. We expected inter-subject variation in physiological reaction to mask the differences in physiological reactions evoked by the presence conditions (e.g., various frame rates). Contrary to this expectation, however, we found strong trends in the physiological measures among conditions in both the Passive Haptics and Frame Rate studies. (The condition was not varied in the Multiple Exposures study.)

In the Passive Haptics study, both ΔHeart Rate and ΔSkin Conductance both varied in the expected direction non-significantly (3.3 BPM, $P = 0.097$; 1.0 mSiemens, $P = 0.137$, respectively).

In the Frame Rate study, ΔHeart Rate followed hypothesized patterns, but ΔSkin Conductance did not. After the anomalous reaction at 10 FPS (as in full data set – compare Figures 6 and 7), ΔHeart Rate differentiated among presence conditions: at 30 FPS it was higher than at 15 FPS, and this difference was nearly significant (7.2 BPM; $P = 0.054$).

Overall, ΔHeart Rate shows promise as a between-subjects measure of presence. Though it did not correlate well with the reported measures (between-subjects), it did differentiate among the conditions with some statistical power in Passive Haptics and Frame Rate. ΔSkin Conductance did not show as much promise as a between-subjects measure. For more discussion of physiological reactions as between-subjects measures of presence, see [Meehan 2001].

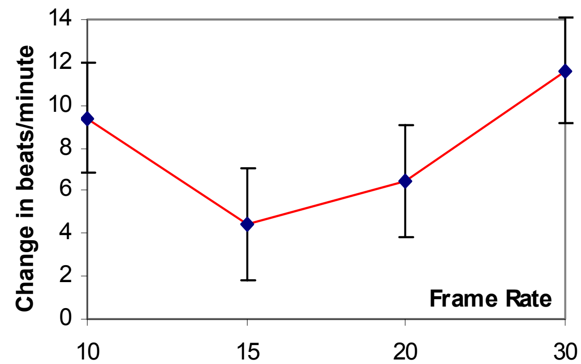

Figure 7. Between-subjects analysis: ΔHeart Rate.

4. VE Effectiveness results

Above we described the experiments as they related to the testing of the physiological presence measures, below we discuss each experiment with respect to the aspect of VEs it investigated.

Effect of Multiple Exposures on Presence. As described in Section 1.4.1, ten users go through the same VE twelve times (over four days) in order to study whether the presence inducing power of a VE declines, or becomes unusably small, over multiple exposures. We did find significant decreases in each presence measure (reported and physiological) in either this experiment or one of the subsequent two experiments (see Table 3). However, none of the measures decreased to zero nor did any become unusably small. The findings support our hypothesis that all presence measures decrease over multiple exposures to the same VE, but not to zero.

Effect of Passive Haptics on Presence. Our hypothesis was that supplementing a visual-aural VE with even rudimentary, low-fidelity passive haptics cues significantly increases presence. This experiment was only one of a set of studies investigating the passive haptics hypothesis. The detailed design, results, and discussion for the set are reported elsewhere [Insko 2001].

We found significant support for the hypothesis in that, with the inclusion of the 1.5-inch ledge, presence as measured by ΔHeart Rate, Reported Behavioral Presence, and ΔSkin Conductance was significantly higher at the $P < 0.05$ level. Reported Presence also had a strong trend ($P < 0.10$) in the same direction.

Effect of Frame Rate on Presence. Our hypothesis was that as frame rate increases from 10, 15, 20, 30 frames/second, presence increases. For frame rates of 15 frames/second and above, the hypothesis was largely confirmed. It was confirmed with statistical significance for 15 to 20 FPS and 15 to 30 FPS. 20 to 30 FPS though not statistically significant was in the same direction. 10 FPS gave anomalous results on all measures except Reported Presence, which increased monotonically with frame rate with no statistical significance.

5. Future Work

Given a compelling VE and a sensitive, quantitative presence measure, the obvious strategy is to degrade quantitative VE quality parameters in order to answer the questions: What makes a VE compelling? What are the combinations of minimum system characteristics to achieve this?

For example, we would like to study the effect of

- Latency
- Self-avatar fidelity
- Aural localization
- Visual Detail
- Lighting Realism
- Realistic physics in interactions with objects
- Interactions with other people or agents

Then we hope to begin to establish trade-offs for presence evoked: Is it more important to have latency below 50 ms or frame rate above 20 FPS?

Additionally, we must eliminate the cables that tether subjects to the monitoring, tracking, and rendering equipment. Our subjects reported this encumbrance as the greatest cause of breaks in presence.

6. Acknowledgements

We would like to thank the University of North Carolina (UNC) Graduate School, the Link Foundation, and the National Institutes of Health National Center for Research Resources (Grant Number P41 RR 02170) for funding this project. We would like to thank the members of the Effective Virtual Environments group, the UNC Computer Science Department, and Dr. McMurray of the UNC Applied Physiology Department. Without their hard work, none of this research would have been possible. We would like to thank Drs. Slater, Usoh, and Steed of the University College of London who built much of the foundation for this work. We would also like to thank the reviewers for their thoughtful comments and suggestions.

7. References

Abelson, J. L. and G. C. Curtis (1989). Cardiac and neuroendocrine responses to exposure therapy in height phobics. *Behavior Research and Therapy,* 27(5): 561-567.

Andreassi, J. L. (1995). *Psychophysiology: Human behavior and physiological response*. Hillsdale, N.J., Lawrence Erlbaum Associates.

Barfield, W., T. Sheridan, D. Zeltzer and M. Slater (1995). Presence and performance within virtual environments. In W. Barfield and T. Furness, Eds., *Virtual environments and advanced interface design.* London, Oxford University Press.

Cowings, P., S. Jensen, D. Bergner and W. Toscano (2001). A lightweight ambulatory physiological monitoring system. NASA Ames, California.

Dillon, C., E. Keogh, J. Freeman and J. Davidoff (2001). Presence: Is your heart in it? *4th Int. Wkshp. on Presence*, Philadelphia.

Ellis, S. R. (1996). Presence of mind: A reaction to Thomas Sheridan's "Further musings on the psychophysics of presence". *Presence: Teleoperators and Virtual Environments*, 5(2): 247-259.

Emmelkamp, P. and M. Felten (1985). The process of exposure in vivo: cognitive and physiological changes during treatment of acrophobia. *Behavior Research and Therapy*, 23(2): 219.

Freeman, J., S. E. Avons, D. Pearson, D. Harrison and N. Lodge (1998). Behavioral realism as a metric of presence. *1st Int. Wkshp. on Presence*.

Guyton, A. C. (1986). Basic characteristics of the sympathetic and parasympathetic function. In *Textbook of Medical Physiology***,** 688-697. Philadelphia, W.B. Saunders Company.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments,* 1: 262-271.

IJsselsteijn, W. A. and H. d. Ridder (1998). Measuring temporal variations in presence. *1st Int. Wkshp. on Presence*.

B. Insko (2001). *Passive haptics significantly enhance virtual environments*, Doctoral Dissertation. Computer Science. University of North Carolina, Chapel Hill, NC, USA.

Kleinbaum, D., L. Kupper, K. Muller and A. Nizam (1998). *Applied regression analysis and other multivariate methods*.

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Brickman and D. J. Rog, Eds., *Handbook of applied social research methods***,** 39-68. Thousand Oaks, California, Sage Publications, Inc.

Lombard, M. and T. Ditton (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication,* 3(2).

McMurray, D. R. (1999). Director of Applied Physiology lab, University of North Carolina. Personal Communication.

M. Meehan (2001). *Physiological reaction as an objective measure of presence in virtual environments*. Doctoral Dissertation. Computer Science. University of North Carolina, Chapel Hill, NC, USA.

Regenbrecht, H. T. and T. W. Schubert (1997). Measuring presence in virtual environments. In Proc. of *Human Computer Interface International*, San Francisco.

SAS (1990). SAS/ STAT User's Guide, Version 6, Fourth Edition. Cary, NC, USA, SAS Institute Inc.

Schubert, T., F. Friedmann and H. Regenbrecht (1999). Embodied presence in virtual environments. In R. Paton and I. Neilson, Eds., *Visual Representations and Interpretations*. London, Springer-Verlag.

Sheridan, T. B. (1996). Further musings on the psychophysics of presence. *Presence: Teleoperators and Virtual Environments,* 5(2): 241-246.

Singleton, R. A., B. C. Straits and M. M. Straits (1993). *Approaches to Social Research*. New York, Oxford University Press.

Slater, M., M. Usoh and A. Steed (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments,* 3(2): 130-144.

Slater, M., M. Usoh and A. Steed (1995). Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer Human Interaction (TOCHI),* 2(3): 201-219.

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments,* 8(5): 560-565.

Slonim, N. B., Ed. (1974). *Environmental Physiology*. Saint Louis. The C. V. Mosby Company.

Sutherland, S. (1996). *The international dictionary of psychology*. New York, The Crossroads Publishing Company.

Usoh, M., K. Arthur, M. Whitton, R. Bastos, A. Steed, M. Slater and F. Brooks (1999). Walking > walking-in-place > flying in virtual environments. In *Proc. of ACM SIGGRAPH 99*. ACM Press/ ACM SIGGRAPH.

Weiderhold, B. K., R. Gervirtz and M. D. Wiederhold (1998). Fear of flying: A case report using virtual reality therapy with physiological monitoring. *CyberPsychology and Behavior,* 1(2): 97-104.

Witmer, B. G. and M. J. Singer (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments,* 7(3): 225-240.