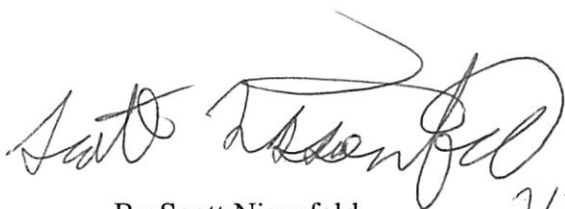



Using Reality to Evaluate the ITC Presence Questionnaire


By Scott Nisenfeld 2/3/03
Department of Computer Science
Brown University

Advisor David Zeltzer
Department of Computer Science
Brown University


2/3/03

Using Reality to Evaluate the ITC Presence Questionnaire

Scott Nisenfeld
Department of Computer Science
Brown University

Abstract

Presence, the sense of “being there”, is a very important idea when discussing the usefulness of Virtual Environments because it is this sense that we want to impart to the person in the virtual environment. It is therefore important to have a way to measure a subject’s sense of presence when in a VE, and there have been many questionnaires created to measure this sense. However, how well do these questionnaires actually measure presence?

This paper follows the procedures put forth by Usoh, Catena, Arman, and Slater to test presence questionnaires, and is an extension of that paper. We carried out a between-group experiment to assess whether the Independent Television Commission Sense of Presence Inventory (ITC-SOPI) can distinguish between real and virtual experiences. We ran a second test to assess if previous exposure to either environment affected the sensation of presence. One group of ten subjects searched for a box in a real library environment two times. A second group of ten subjects first searched for a box in the real environment and then in a virtual environment that simulated the same library. A third group of ten subjects searched the virtual environment of the library two times. A fourth group of ten subjects searched first the virtual environment and then the real library environment. Immediately after each search task was completed, subjects were given ITC-SOPI by Lessiter, Freeman, Keogh, and Davidoff (Lessiter, Freeman, Keogh, E., and Davidoff, 2001). This paper follows the procedures put forth by Usoh, Catena, Arman, and Slater to test presence questionnaires. That paper argues that a presence questionnaire should be able to pass a “reality test,” whereby under current conditions the four “presence” factors generated from a real experience should be distinguishable from four “presence” factors generated from a similar virtual experience. The ITC-SOPI questionnaire was statistically able to differentiate between both environments for all factors except for mean *engagement*. Furthermore, the ITC-SOPI shows that the order in which the environments were encountered does not influence the reported four factors. We concluded that this questionnaire does pass the “reality test”, and is useful for comparisons of experience across presentation methods, such as between immersive virtual environments compared to real environments, or a desktop presented environment compared to immersive virtual environment.

1 Introduction

The concept of *presence* is very important in the study of virtual environments. One might argue that the purpose of a VE is to fool the subject into thinking that what they experience in the VE is realistic. Papers discussing presence (Held & Durlach, 1992; Loomis, 1992; Sheridan, 1992; Zeltzer, 1992; Heeter, 1992; Steuer, 1992) have suggested many factors of presence. These factors were distilled in the Usoh et al. (2000) to be

1. High-resolution information displayed to the participant in a manner that does not indicate the existence of the display devices. This includes Steuer's notion of vividness, "the ability of a technology to produce a sensorially rich mediated environment"
2. Consistency of the displayed environment across all sensory modalities.
3. The possibility of the individual being able to navigate through – and interact with – objects in the environment, including interaction with other actors who may spontaneously react to the individual;
4. The individual's virtual body, their self-representation within the environment, should be similar in appearance or functionality to the individual's own body, and respond appropriately to the movement of their head, eyes, and limbs.
5. The connection between an individual's actions and the effects of those actions should be simple enough for the individual to quickly learn.

It is obvious that these factors are at their acme for a healthy individual in the context of everyday reality.

Usoh et al. (2000) go on to suggest that if presence is optimal for real-world experiences, methods used to measure presence should be able to distinguish between a real environment and a VE mock up. Many papers have used questionnaires to grade VEs, but Usoh et al (2000) suggest that these methods be first put through a "reality test." They stress that it is implicit "that presence should therefore be higher in a real than in a VE, other things being equal."

In this paper, we use this reality test protocol to grade the ITC-SOPI proposed by Lessiter, Freeman, Keogh, and Davidoff (2001). The ITC-SOPI was created using an iterative process of proposing questions, implementing the questionnaire with different media, scoring the results, and then refining the questions. The questionnaire was created from scoring media as varied as cinema film shown in a student union, to a video game, to the IMAX 3D format. They, however, did not use an interactive immersive VE setup, or a real environment. The questionnaire grades on a scale of 1 thru 5, with ten questions about feelings after the experience and thirty-eight questions about feelings experienced while in the environment. We receive four scores that measure Spatial Presence, Engagement, Naturalness, and Negative Effects, with 5 being extreme agree and 1 being extreme disagree. Lessiter et al define each group as such

- Spatial Presence – sense of physical placement in mediated environment, interaction with, and control over parts of the mediated environment.
- Engagement – sense of being psychologically involved and enjoying the content presented.
- Naturalness – sense that the mediated environment is lifelike and real.
- Negative Effects – sense of adverse psychological reactions to environment presented.

While Spatial Presence encompasses the given definition of *presence*, the other measures relate to the presence factors one, two and five. These measures equal the mean scores from select questions, thus providing each measure with a final score between one and five.

2 Method

The current study involved forty people searching for a red box hidden on a floor of Brown University's medical library (called the "library"). There were two version of this search space, the actual library and a VRML mock up displayed in a CAVE (see Figure 1 and 2). The participants where recruited by advertisements posted on the campus. They where then assigned randomly to one of four groups. Before each test the subject was placed in a small 4ft by 4ft room (the elevator for real environment, a simulated room for VE) to allow participants to grow accustomed to navigating in the VE world. Group A was run though the real library twice, group B ran though first the library then the VE, group C ran though first the VE then the library, and group D ran though the VE twice. Each subject received a break of about ten minutes between each run through the environment. The group broke down as follows; Group A, Male: 4 Female: 6; Group B, Male: 6 Female: 4; Group C, Male: 7 Female: 3; Group D, Male: 3 Female: 7. All but 3 of the forty participants where students, the others being either staff at the university, or recently graduated students. For simplicity we will now refer the groups to as such; Group A is Real/Real, Group B is Real/VE, Group C is VE/Real, and Group D is VE/VE. All subjects where paid \$10 for their time.

The average time to search in the real environment was about three minutes, in the VE it was about five minutes. At the end of each search, the subject answered the questionnaire, and then they either ran the same environment again or switched environments and did the questionnaire again. The ITC-SOPI questions in relation to "displayed environment" where put into the context of what was viewed for the REAL group.

The scenarios were implemented on an Onyx 2 with 8 CPUs running at 200 MHz each and 2 Gigs of Ram. The motion tracker was a Polyhemus Magnetic Tracker, with a latency of approximately 33-34 milliseconds. The total scene consisted of roughly 3,000 polygons displayed no less then 20 Hz in stereo. The latency was approximately 33-34 ms.

Subjects moved through the environment at a constant velocity by pressing a thumb button on a 3D mouse, they could either turn physically or rotate the scene via the 3D mouse thumb stick.

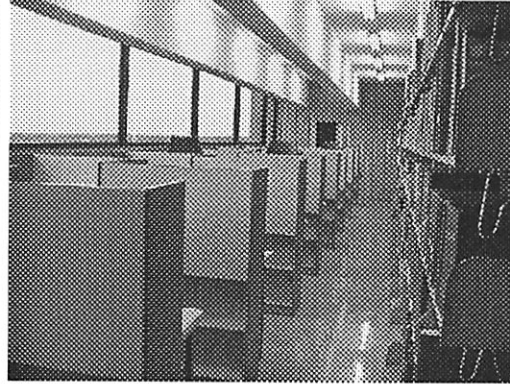
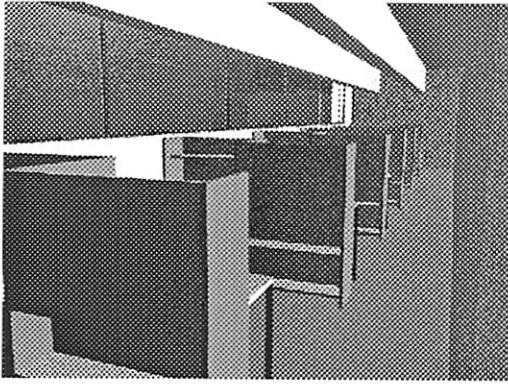


Figure 1. A view of the virtual Library space. **Figure 2.** A view of the real Library space.

3 Results

The fundamental result is that the ITC-Sense of Presence Inventory was able to distinguish between both the real and virtual environments in three out of four of its measures. We took the scores from each of the groups first run, creating two groups of twenty each, real and VE, to see if there was a significant difference between the scores. For all dependent measures except Engagement the differences were significant at the $p = .05$ level. This could be attributed to the fact that the user did not have to touch anything in either environment, and as such reported a similar score to the VE.

We began by looking at only the first run of the experiment for each subject to determine if the questionnaire passed the reality test. Thus, the groups broke down into two groups with 20 subjects each, a group of subjects that ran in the Real environment and one that ran in the VE. Using a two-sample t-test, we compared the two groups. Significant differences ($p < .05$) were found for all dependent measures except Engagement ($p < .10$). The questions that the questionnaire uses to measure the task and the simplicity of the task account for this aberration. Questions such as feeling sad about the ending of the experience and another about wishing the experience to continue mean two different things depending on the environment. Many in the real environment voiced "feeling frustrated" when they did not find the object quickly, while those in the VE voiced feeling that they wish that they had more time to explore the VE completely, even after finding the object. Also, given the short time that subjects were able to locate the object, questions relating to time seem to lose their effectiveness in accurately gauging Engagement.

Table 1 Means and Standard Deviations of ITC Scores on First Trial

First Trial	Mean Spatial Presence T1	Mean Engagement T1	Mean Naturalness T1	Mean Negative Effects T1
Real ($n = 20$)	$3.93 \pm .70$	$3.31 \pm .55$	$4.42 \pm .63$	$1.66 \pm .72$
Virtual ($n = 20$)	$3.10 \pm .60$	$3.57 \pm .45$	$2.96 \pm .82$	2.57 ± 1.10

Table 2 Means and Standard Deviations of ITC Scores on Same/Different Environments Experienced

Group	Mean Spatial Presence T1	Mean Engagement T1	Mean Naturalness T1	Mean Negative Effects T1
Same ($n = 20$)	3.55 \pm .84	3.36 \pm .68	3.62 \pm 1.08	2.35 \pm 1.46
Different ($n = 20$)	3.72 \pm .78	3.10 \pm .57	3.93 \pm 1.09	2.20 \pm 1.03

Therefore, now that we have shown that the questionnaire is reliable in the cases of Spatial Presence, Naturalness, and Negative Effects, does a previous experience in a similar/same environment have a noticeable effect on a subject's sense of presence? First, we used a two sample t-test to see if the subject's sense of presence was affected in the second session based on if it was the same or different from the first session. As shown in Table 2, some of the measures are higher (Spatial Presence and Naturalness) when the subject experienced a different environment. However, this perceived higher value is not statistically significant ($p \approx .5$). If there was a statistical significant difference, then a conclusion could be drawn that a change in environments has some measurable impact on presence.

Next, we calculated if the difference in reported presence between the two sessions differs for the four groups. We ran a one-way ANOVA test and a post-hoc Tukey HSD over the data to determine which groups differed from the others. Table 3 shows the means and deviation for each groups' difference of Trial One from Trial Two. Here we see that for all groups, except group Real/VE, Spatial Presence was higher in the second run, with group VE/Real having the greatest positive difference. Looking at the trend for all measures, we see that group Real/Real and VE/VE are consistently lumped together into the same subset by the Tukey HSD test. Ignoring Engagement, we see in Table 4-7 that group VE/Real's and Real/VE's scores difference are positive and negative as one would expect with the change of Environment.

Table 3 Means and Std. Deviation for the Difference in Measures between Trial One and Trial Two

	Group	N	Mean	Std. Deviation	Std. Error
Mean Spatial Presence Diff	Real/Real	10	-.01	.19	.06
	Real/VE	10	.92	.89	.28
	VE/Real	10	-1.37	.68	.21
	VE/VE	10	-.01	.30	.10
Mean Engagement Diff	Real/Real	10	.02	.49	.15
	Real/VE	10	-.35	.64	.20
	VE/Real	10	.95	.53	.17
	VE/VE	10	.23	.34	.11
Mean Naturalness Diff	Real/Real	10	.03	.31	.10
	Real/VE	10	1.60	.93	.29
	VE/Real	10	-2.06	.83	.26
	VE/VE	10	.08	.44	.14
Mean Negative Effects Diff	Real/Real	10	.15	1.16	.37

Real/VE	10	-1.25	1.03	.33
VE/Real	10	.43	.72	.23
VE/VE	10	.02	.70	.22

Table 4 Tukey HSD Mean Spatial Presence Diff
Means for groups in homogeneous subsets are displayed.

Group	N	Subset for alpha = .05		
		1	2	3
VE/Real	10	-1.37		
Real/Real	10		-.01	
VE/VE	10		-.01	
Real/VE	10			.92
Sig.		1.00	1.00	1.00

Table 5 Tukey HSD Mean Engagement Diff
Means for groups in homogeneous subsets are displayed.

Group	N	Subset for alpha = .05	
		1	2
Real/VE	10	-.35	
Real/Real	10	.02	
VE/VE	10	.23	
VE/Real	10		.95
Sig.		.07	1.00

Table 6 Tukey HSD Mean Naturalness Diff
Means for groups in homogeneous subsets are displayed.

Group	N	Subset for alpha = .05		
		1	2	3
VE/Real	10	-2.06		
Real/Real	10		.03	
VE/VE	10		.08	
Real/VE	10			1.60
Sig.		1.00	.99	1.00

Table 7 Tukey HSD Mean Negative Effects Diff
Means for groups in homogeneous subsets are displayed.

Group	N	Subset for alpha = .05	
		1	2
Real/VE	10	-1.25	
VE/VE	10		.02
Real/Real	10		.15
VE/Real	10		.43
Sig.		1.00	.75

Given these two statistical tests we ran, it is easy to see that while the scores did not change much when the same environment was encountered, the fluctuations of scores within groups VE/Real and Real/VE can be seen to be connected more to the change in environment than any rise in overall presence related to previous experience.

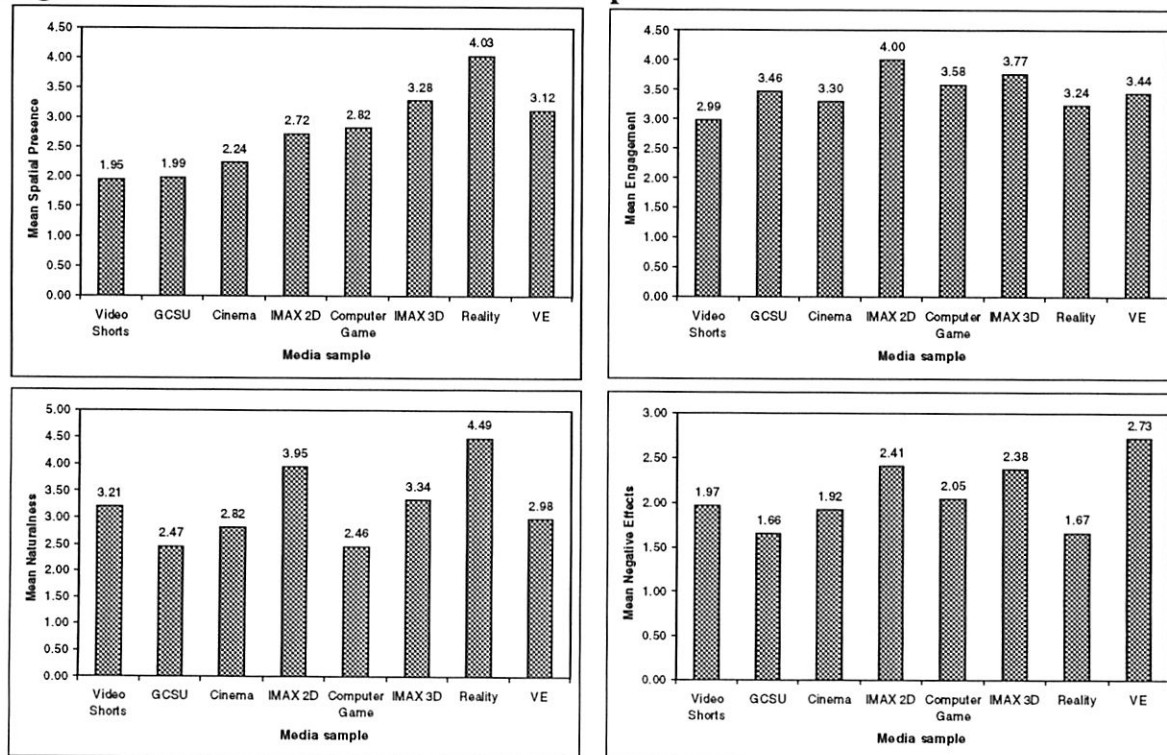
As a final test, we ran a t-test on the second trial to see if the questionnaire could pass the reality test regardless of previous experience. Table 8 shows the means and deviations. As with the t-test from the first trial, all had two-tailed Sig. values of .000 to .001, except for Engagement. This again strengthens assumptions that the questionnaire can reliably differentiate between reality and a VE mockup.

Table 8 Means and Standard Deviations of ITC Scores on Second Trial

Second Trial	Mean Spatial Presence T2	Mean Engagement T2	Mean Naturalness T2	Mean Negative Effects T2
Real ($n = 20$)	4.13 \pm .68	3.16 \pm .70	4.55 \pm .61	1.68 \pm .91
Virtual ($n = 20$)	3.14 \pm .60	3.30 \pm .56	3.00 \pm .88	2.88 \pm 1.27

If we insert the acquired mean scores into a graph of scores provided in Lessiter et al. (figure 3), we see that the VE falls reliably within the other media types. Reality is consistently on the upper or lower bounds (except for Engagement, which we discussed previously). It is interesting to note that the score for the real environment's negative effects seems high. We speculate that this has to do with the subject coming into the experiment with such effects, rather than the environment causing these adverse effects. Still it is an interesting fact to take notice of for further refinement of the questionnaire.

Figure 2 Mean Scores for Each Media Sample



4 Conclusions

This paper has evaluated the ITC-Sense of Presence Inventory, using the setup put forth in Usoh et al. (2000). The questionnaire was able to differentiate between a reality setting and a VE setting, two media presentations that were not used in the statistical creation of the questionnaire. In referring the mean scores to those presented in Lessiter et al. (2001), we find that the real environment scored highest in Spatial Presence and Naturalness and equal to watching a VHS film in terms of negative effects. The VE lays in-between the computer game and IMAX3D for Spatial Presence, relatively equal to a video game for engagement, between video shorts and cinema for naturalness, and slightly higher to IMAX2D for negative effects.

The only part where it failed was in differentiating between the engagement of the real world and VE, with the VE scoring higher. We speculate that this might have to do with the fact that the subjects did not have to interact with either environment, and as such, many did not attempt to interact with the real environment. Given the nature of the VE environment, physical interactions between subject and environment were limited to using the 3D mouse for navigation. The need to interact with the 3D mouse to move can account for the slightly higher engagement scores in the VE. As such, a lack of interaction built into the experiment as opposed to an overall deficiency in the questionnaire explains this aberration.

We find that previous exposure to an environment, either real or virtual, has no statistical impact on the four measures. Perhaps the rapid succession that the environments were presented caused the participant to contrast the two rather than judge both on their own merits. Whatever the case, a more comprehensive study would yield better results.

It would be imprudent to suggest that from this one experiment that the ITC-SOPI should be made a standard or that it even is the best questionnaire available. These results provide an interesting starting place for further testing, such as applying the questionnaire (and others) on a more environmentally interactive type of task. Since during the creation of the ITC-SOPI neither a real environment nor a high-grade VE were used, the data collected can be used to further strengthen the questionnaire. Considering that most of the media used to build the questionnaire were non-interactive, nor employ 3D displays (save for the computer game and the 3D IMAX), the questionnaire performed reliably and predictably. The comparison of the means created here to those presented by Lessiter et al. (2001) shows a reliable and predictable scoring. With more use, more feedback, and further tweaking the ITC could become a powerful tool.

Note: The ITC-Sense of Presence Inventory (ITC-SOPI) is copyright of the UK Independent Television Commission.

References

- Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1(2), 262-271.
- Held, R.M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments*, 1, 109-112.
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), 282-297.
- Loomis, J.M. (1992). Distal attribution and presence. *Presence: Teleoperators and Virtual Environments*, 1, 113-119.
- Sheridan, T.B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1, 120-126.
- Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4), 73-93.
- Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5), 497-503.
- Zeltzer, D. (1992). Autonomy, interaction, and presence. *Presence: Teleoperators and Virtual Environments*, 1, 127-132.