Martin Usoh
Ernest Catena
Sima Arman
Mel Slater
Department of Computer Science
University College London
London WC1E 6BT

Using Presence Questionnaires in Reality

Abstract

A between-group experiment was carried out to assess whether two different presence questionnaires can distinguish between real and virtual experiences. One group of ten subjects searched for a box in a real office environment. A second group of ten subjects carried out the same task in a virtual environment that simulated the same office. Immediately after their experience, subjects were given two different presence questionnaires in randomized order: the Witmer and Singer Presence (WS), and the questionnaire developed by Slater, Usoh, and Steed (SUS). The paper argues that questionnaires should be able to pass a "reality test," whereby under current conditions the presence scores should be higher for real experiences than for virtual ones. Nevertheless, only the SUS had a marginally higher mean score for the real compared to the virtual, and there was no significant difference at all between the WS mean scores. It is concluded that, although such questionnaires may be useful when all subjects experience the same type of environment, their utility is doubtful for the comparison of experiences across environments, such as immersive virtual compared to real, or desktop compared to immersive virtual.

1 Introduction

The concept of presence in a virtual environment (VE) naturally raises the question of the meaning of *presence* with respect to real-world experiences. (Let's leave aside here ontological questions about the nature of existence.) A normally healthy, able, conscious individual acting in a physical environment is fully present in that environment with respect to the meaning of presence in a VE. Consider, for example, the early papers discussing the notion of presence (Held & Durlach, 1992; Loomis, 1992; Sheridan, 1992; Zeltzer, 1992; Heeter, 1992; Steuer, 1992). A summary of the factors thought to influence presence as introduced in these papers include

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

Presence, *Vol. 9, No. 5, October 2000,* 497–503

own body, and respond appropriately to the movements 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.

Considering each one of these, of course they are at their optimum in the context of a person acting in everyday reality.

Taking into account more-recent discussions of presence, Bystrom, Barfield, and Hendrix (1999) introduce the Immersion, Presence, and Performance model of interaction in VEs. It is implicit in their approach to presence that this would be optimal for people in the real world as compared to a virtual world. Following a quite different perspective, Zahoric and Jenison (1998) give a definition of presence as ''tantamount to successfully supported action in the environment.'' Their approach argues that reality is grounded in action rather than in mental filters and that ''the reality of experience is defined relative to functionality, rather than to appearances'' (Flach & Holden, 1998). This approach to presence is concerned with action rather than the appearance of how things look and sound. In other words ''being there'' is the ability to act there. It is clear that, in this approach also, the greatest potential for people to perform successful action is in the real world. (Whether or not they can actually carry out these actions because of questions of personal ability or power relationships is another matter.) A thorough review of the literature on presence can be found in Draper, Kaber, and Usher (1998).

If presence is optimal for real-world experiences, methods that attempt to elicit or measure presence should be able to discriminate between experiences that take place in a physical environment and virtual environment. The vast majority of papers that have studied presence use questionnaires for this purpose. We propose that any method for measuring presence, most especially questionnaires, should be subject to a ''reality test.'' In other words, they should be used for a group of subjects acting in a real environment, and the results compared with a group of subjects acting in a virtual environment. Other things being equal, a measure passes the reality test only if the measured presence is greater for the real environment, under today's conditions, and for the foreseeable future.

A Turing test in artificial intelligence discriminates the responses of a computer program from those of a human being. If an observer cannot distinguish between these, then the computer program has passed the Turing test and can be thought of as a thinking machine. The idea of a sort of Turing test for VEs has been proposed in the literature. For example, Steuer (1992) suggested that a methodology for measuring the efficacy of a virtual environment would be the extent to which subjects could not discriminate between it and a real environment. The approach of Schloerb (1995) towards measuring presence is along similar lines. It is implicit in these suggestions—and perhaps so obvious that it is overlooked—that presence should therefore be higher in a real than in a VE, other things being equal.

In this paper, we report an experiment that had the sole purpose of subjecting two presence questionnaires to this reality test. The first is a formally proposed and evaluated questionnaire constructed by Witmer and Singer (1998). Their presence questionnaire consists of 32 questions each on a 1-to-7 scale. Each question corresponds to a factor thought to be correlated with presence, such as, for example: ''How much were you able to control events?'' The final presence score is the sum of 32 such questions. We refer to this questionnaire as WS.

The second approach is that developed over a number of studies by Slater and colleagues and most recently used in Slater et al. (1998) and Usoh et al. (1999). This questionnaire is based on several questions that are all variations on one of three themes: the 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.'' Suppose, in a particular study, that n such questions are asked, each on a 1-to-7 scale with the higher score indicating greater presence. The presence score is taken as the number of answers r out of n that have a score of 6 or 7. In particular, if this measure of presence is to be used as a response variable in a regression analysis on a number of explanatory and independent variables, then logistic regression would be em-

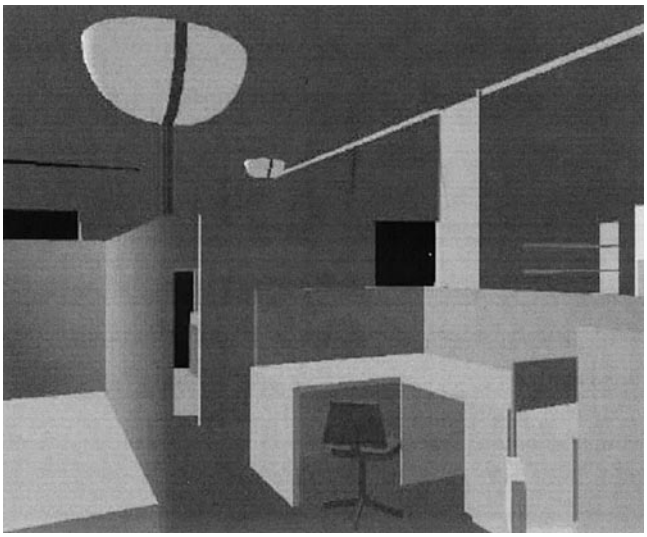

Figure 1. A view of the virtual office space.

Figure 2. A view of the real office space.

ployed rather than normal regression. We refer to this questionnaire as SUS (Slater-Usoh-Steed).

2 Method

The study involved twenty people searching for a red box hidden in a university research laboratory (called the ''office space'' from now on). There were two versions of this space: the real space and a virtual environment model of this space. (See Figure 1 and 2). The participants were recruited by advertisement on the university campus, and were assigned randomly to one of two groups. The first group (V) experienced only the virtual office space, and carried out their task of searching for the box entirely within this VE. The second group (R) carried out the same task in the real office space. All of the subjects were students. In the V group were three women and seven men, in the R group two women and eight men. The subjects were paid the equivalent of $9.

The search task took between seven and fourteen minutes for those in the virtual office, and between six and ten minutes for those in the real office. At the end of the search task, the subjects answered the two questionnaires (WS and SUS). To avoid the possibility that the order of answering the questionnaires could make any difference, this was randomized: half of the subjects in each group first answered WS and then SUS, and the other half in the reverse order.

Almost all of the questions in the WS questionnaire make sense in relation to a real environment. This questionnaire was used as published in the original article. The SUS questionnaire as used in the above-cited recent studies was employed, the only change being that direct references to an experience within a VE were changed to refer to the ''office space.'' Hence, the same questionnaire could be used for both V and R groups. The six questions relating to presence were as follows.

1. Please rate *your sense of being in the* office space, on the following scale from 1 to 7, where 7 represents your *normal experience of being in a place.*
   I had a sense of ''being there'' in the office space:
   (1) Not at all. (7) Very much.
2. To what extent were there times during the experience when the office space was the reality for you?
   There were times during the experience when the office space was the reality for me. . .
   (1) At no time. (7) Almost all the time.
3. When you think back about your experience, do you think of the office space more as *images that you saw,* or more as *somewhere that you visited?*
   The office space seems to me to be more like. . .

(1) Images that I saw. (7) Somewhere that I visited.

4. During the time of the experience, which was strongest on the whole, your sense of being in the office space, or of being elsewhere?
   I had a stronger sense of. . .
   (1) Being elsewhere. (7) Being in the office space.
5. Consider your memory of being in the office space. How similar in terms of the *structure of the memory* is this to the structure of the memory of other *places* you have been today? By ''structure of the memory,'' consider things like the extent to which you have a visual memory of the office space, whether that memory is in color, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such *structural* elements.
   I think of the office space as a place in a way similar to other places that I've been today. . .
   (1) Not at all. (7) Very much so.
6. During the time of the experience, did you often think to yourself that you were actually in the office space?
   During the experience I often thought that I was really standing in the office space. . .
   (1) Not very often. (7) Very much so.

The scenarios were implemented on an SGI Onyx with twin 196MHz R10000, Infinite Reality Graphics and 192M main memory. The tracking system has two Polhemus FASTRAKS, one for the HMD and another for a five-button 3-D mouse. The helmet was a Virtual Research VR4 with a resolution of 742 × 230 pixels for each eye, 170,660 color elements, and a field of view of 67 deg. diagonal at 85% overlap.

The total scene consisted of 12,564 polygons which ran at a frame rate of no less than 20 Hz in stereo. The latency was approximately 120 ms.

Subjects moved through the environment in gaze direction at constant velocity by pressing a thumb button on the 3-D mouse.

Table 1. *Means and Standard Deviations of Questionnaire Scores*

Group	WS	SUS count	SUS mean
Virtual ($n = 10$)	90.3 ± 14.5	1.0 ± 1.7	3.8 ± 1.3
Real ($n = 10$)	90.6 ± 18.4	2.0 ± 2.0	4.4 ± 1.5

3 Results

The fundamental result is that neither WS nor SUS distinguished between the real and virtual experiences. The results are shown in Table 1. The SUS Count column shows the mean of the SUS count of 6 or 7 scores among the six questions. The SUS Mean column uses the mean score across the six questions instead. Clearly, there are no significant differences between the groups for the WS or SUS Mean columns. The order in which the questionnaires were presented likewise made no difference to this result.

Treating the SUS Count response as binomially distributed for a logistic regression on group, as intended for the SUS approach, there is a significant difference between the groups ($\chi^2 = 4.511$ on 1 d.f. is significant at 5%). This difference is entirely accounted for by two of the six questions (3 and 6). Hence, there is some marginal evidence that the SUS questionnaire satisfies the reality test, but, when the mean results for the individual questions are examined (Table 2), not much confidence can be put in this.

Consider, for example, question 1. The mean score for those in the real office space group was only 4 (in response to a question about a sense of really being in the office space)!

The SUS questionnaire had a free-response final question:

> Please write down any further comments that you wish to make about your experience. In particular, what things helped to give you a sense of ''really being'' in the office space, and what things acted to ''pull you out'' of this?

Table 2. *Means and Standard Deviations of Responses to the Individual SUS Questions*

Group	Q1	Q2	Q3	Q4	Q5	Q6
Virtual	4.3 ± 1.5	3.6 ± 1.3	2.6 ± 1.6	4.6 ± 1.3	4.4 ± 2.3	3.3 ± 2.2
Real	4.0 ± 2.1	4.0 ± 2.3	4.6 ± 2.2	5.1 ± 1.9	3.7 ± 1.9	5.2 ± 1.6

For those in the virtual group, the responses were as expected for a VR experience. Activity such as bending down and looking underneath tables tended to reinforce the sense of being in the office space. The lag in response to head movements, the lack of collision detection with virtual objects, and interference from physical cables and bumping into other real objects all tended to bring subjects out of the experience.

Of greater interest are the comments of those in the real office space, for these give some insight into how they interpreted the overall questionnaire. Reasons given for the enhancement of being in the office space were

1. The office being similar to another one in which the subject had previously worked.
2. Sounds such as people talking on the phone.
3. Visual cues such as computers and paper clips.
4. Natural lighting.
5. Other people's reaction to his presence gave a greater sense of being in the office space. ''It made you aware of your own presence in a way.''
6. Being given the task of having to look for something in the office.
7. The layout, the way that the individual desks and PCs were arranged.
8. Heightened sensitivity caused by being in a new environment.
9. ''Only felt very much in the office when someone turned and asked if they could help.''

The reasons given for being pulled out of the office space were

1. The reactions of ''strangers'' in the office space, who did not know the subject nor why she was there.
2. The lack of any smell.
3. The task itself not involving much physical interaction or speech.
4. Being ignored by other people.
5. The fact that not everyone were at their desks and the fact that there was not much noise.
6. Lack of familiarity with the office space.

An interpretation is that those in the virtual office relate their sense of presence more to the system factors: how ''real'' is the experience in the sense of the responsiveness of the environment to the individual's actions (for example, you can really bend down and look underneath objects), but, on the other hand, there is a delay in response to head movements. Moreover, the real world occasionally intrudes (the cables). For those in the real office, there is also some degree of comparison with an otherwise normal or perhaps idealized real environment (visual, auditory, and even olfactory issues were mentioned), but also ''being there'' is at least partially reinterpreted as the degree of involvement or lack of alienation in the environment.

There was an interesting relationship between the WS and SUS scores. Figure 3 shows the scatterplot of WS against SUS scores for the group who experienced the virtual office. The correlation was not significant ($r^2 = 0.22$), largely because low SUS scores correspond to a wider range of WS scores. Figure 4 shows the plot for those who experienced the real office. In this case, the correlation is significant at 5% ($r^2 = 0.68$). It seems that high SUS scores are consistently associated with high WS scores, but this is not the case for low SUS scores.

Finally, included in the SUS questionnaire was a question relating to task performance:

> Overall, how well do you think that you achieved your task?
> *I achieved my task. . .*
> (1) Not very well at all. (7) Very well.

This was uncorrelated with both SUS and WS scores for

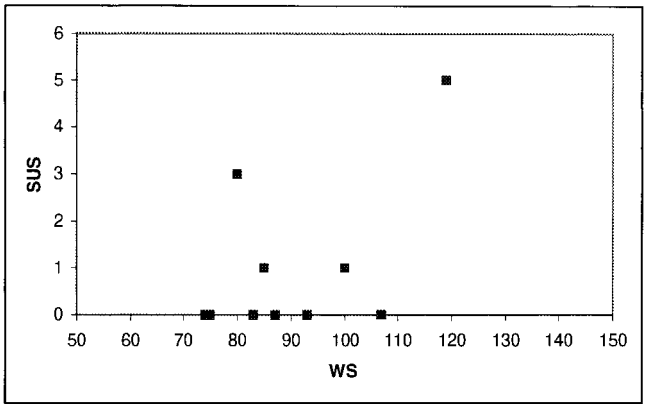

Figure 3. *Scatterplot of WS against SUS for the virtual group.*

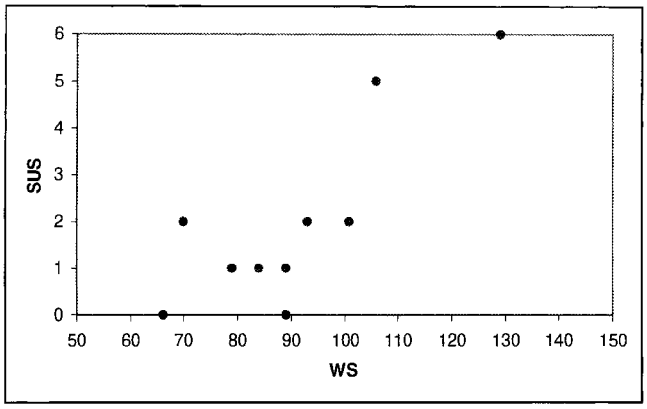

Figure 4. *Scatterplot of WS against SUS for the real group.*

each of the real and virtual environments. Moreover, each person's time to completion was recorded, and again this was uncorrelated with the presence scores.

4 Conclusions

This paper has suggested a reality test for questionnaires designed to measure presence. The test is that the questionnaire should be applied to individuals in a real environment, and should consistently result in higher ''presence'' responses than for people in a virtual environment, other things being equal. This test was applied to two presence questionnaires. The Witmer and Singer questionnaire showed no significant differences in measured presence between a group carrying out a search task in a virtual environment and another group carrying out the task in a corresponding real environment. The Slater-Usoh-Steed questionnaire did show a small, statistically significant difference between the real and virtual results in the expected direction, although all but two of the individual components of the questionnaire showed no difference. The power of the experimental design is relatively low, so a further, larger-scale study is needed. However, the expected differences between the real and the virtual are so large (compare Figures 1 and 2—the real world is so much richer than the virtual), that very large differences between the groups would have been expected, which even a low-power design should have detected. Problems of instability in subjective questionnaires for presence have been recently discussed also by Freeman et al. (1999), although from quite a different viewpoint, examining the influence of prior experience.

An important interpretation of the result is that, if a scientist asks a lay person a question, for example, concerning their ''sense of being there,'' that person will come up with some interpretation that makes the question seem sensible, and then answer that question. If someone is asked for their sense of ''being there'' on a 1-to-7 scale, it gives them permission to answer with a score of less than 7 even when they are really there. The questions are reinterpreted to make sense in the given context. In a VE, *presence* is interpreted as the sense of being in the environment that is depicted by the computer-generated displays and the ability to act in that environment. In the real world, because there is no doubt that the individual is present in the obvious sense, it becomes reinterpreted as the sense of involvement, the lack of isolation, perhaps even the degree of comfort. The thought ''I am not comfortable being here'' might lead to a low presence response.

We speculate that subjects in an experiment will relativize their responses to presence questions to the domain of their given experimental experiences only. For example, it was recently found that, among thirty subjects all experiencing the same shared VE, those who used an immersive head-tracked, head-mounted display system did not report a higher overall sense of presence than those using a desktop system, using a variant of the

SUS questionnaire (Slater et al., 2000). On the other hand, when all subjects experience the same type of immersion, there are consistent variations of presence with other control factors such as an interaction technique (Usoh et al., 1999). The problem is ''cross-environment'' comparisons (virtual to real, immersive to desktop), which do not seem to be valid using this approach.

Acknowledgments

This research was performed as part of the Masters course on Vision, Imaging, and Virtual Environments in the Department of Computer Science at University College London.

References

Bystrom, K-E., Barfield, W., & Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 8*(2), 241–244.

Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors, 40*(3), 354–375.

Flach, J. M., & Holden, J. G. (1998). The reality of experience: Gibson's way. *Presence: Teleoperators and virtual environments, 7*(1), 90–95.

Freeman, J., Avons, S. E., Pearson, D. E., & IJsselstijn, W. A. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments, 8*(1), 1–13.

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.

Loomis, J. M. (1992). Distal attribution and presence. *Presence: Teleoperators and Virtual Environments, 1,* 113–119.

Schloerb, D. (1995). A quantitative measure of telepresence. *Presence: Teleoperators and Virtual Environments, 4*(1), 64–80.

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments, 1,* 120–126.

Slater, M., Steed, A., McCarthy, J., & Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Factors, 40*(3), 469–477.

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in virtual and real environments: A comparative study. *Presence: Teleoperators and Virtual Environments, 9*(1), 37–51.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73–93.

Usoh, M., Arthur, K., Whitton, M., Bastos, R., Steed, A., Slater, M., & Brooks, F. (1999). Walking > Walking-in-Place > Flying in Virtual Environments. *Computer Graphics* (SIGGRAPH), Annual Conference Series, 359–364.

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

Zahoric, P., & Jenison, R. L. (1998). Presence and being-in-the-world. *Presence: Teleoperators and Virtual Environments, 7*(1), 78–89.

Zeltzer, D. (1992). Autonomy, interaction, and presence. *Presence: Teleoperators and Virtual Environments, 1,* 127–132.