

Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire

Mel Slater

Witmer and Singer recently published a questionnaire for eliciting presence in virtual environments together with a questionnaire for measuring a person's immersive tendencies (Witmer & Singer, 1998). The authors mentioned that they did not agree with my notion of immersion: 'Though the VE equipment is instrumental in enabling immersion, we do not agree with Slater's view that immersion is an objective description of the VE technology'. On first reading I was happy to take this as simply a difference of terminology - which is what it is. I had *defined* the term *immersion* to mean the extent to which the actual system delivers a surrounding environment, one which shuts out sensations from the 'real world', which accommodates many sensory modalities, has rich representational capability, and so on (described, for example, in Slater & Wilbur, 1997). These are obviously measurable aspects of a VE system. For example, given two VE systems, and other things being equal, if one allows the participant to turn their head in any direction at all and still receive visual information only from within the VE then this is called (in my definition) a more 'immersive' system than one where the participant can only see VE visual signals along one fixed direction. Given two systems, if one has a larger field of view than the other, then the first is (in my definition) more immersive than the second. As a last example, if one generates shadows in real-time and the other does not, then again, the first is called (by me) more immersive. These are examples of what I mean by more or less 'immersion'. Clearly for all of these types of things metrics can be established which are descriptions of the system, and not descriptions of people's responses to the system. Witmer and Singer, however, define immersion as the person's *response* to the VE system. This difference in terminology is unfortunate, but not a matter of any great concern. In order to

Measuring Presence

avoid further confusion I will use the term ‘system immersion’ to denote my meaning of the word, and ‘immersive response’ to denote W&S’s meaning.

On a second reading of their paper, I realised that this terminological difference was a sign of a very profound difference in methodology regarding the elicitation of presence. The purpose of this note is to explain why I would never use the W&S questionnaire for studying presence - even though I am sure that in itself it can lead to useful insights about the nature of VE experience.

About the meaning of presence itself, I find W&S’s various summary descriptions of the concept helpful:

- Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another.
- ...presence refers to experiencing the computer-generated environment rather than the actual physical locale.

I would also include W&S’s notion of immersion as part of my understanding of the meaning of *presence*:

- Immersion is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences.

In several papers (see references cited in Slater & Wilbur, 1997) I have employed the notion that presence includes three aspects:

- The sense of ‘being there’ in the environment depicted by the VE.
- The extent to which the VE becomes the dominant one - i.e., that participants will tend to respond to events in the VE rather than in the ‘real world’.
- The extent to which participants, after the VE experience, remember it as having visited a ‘place’ rather than just having seen images generated by a computer.

The first is part of the accepted notion of presence (for example, Held & Durlach, 1992, Sheridan, 1992). The second two have come from observing and listening to participants themselves in many

experimental studies. The second, for example, includes the looming response - the participants know that there's 'nothing there' but they still duck when an object flies towards them. The visual cliff experience is another example - where subjects know that they are not going to fall, but nevertheless exhibit signs and symptoms of fear of heights. Another example is where subjects incorporate externally generated sounds from outside the VE into their subjective VE experience, much like physical events are incorporated into ongoing dreams, as noted by Freud. The third category also comes from listening to subjects - when they report that they'd had an experience of being in a *place*, just like any other place they had been earlier in the day. This 'experiencing-as-a-place' is very much what I have tried to convey as a meaning of presence in VEs: people are 'there', they respond to what is 'there', and they remember it as a 'place'. If during the VE experience it were possible to ask the question 'where are you?' - an answer describing the virtual place would be a sign of presence. However, this question cannot be asked - without itself raising the contradiction between where they know themselves to be and the virtual place that their real senses are experiencing.

Whatever the precise meaning of presence a valid problem for those studying the field is to ask: How can presence be 'maximised'? What characteristics does a system have to have in order to increase the likelihood that an individual will experience a high degree of presence? Note that by 'system' I do not simply mean the hardware and system software, but also the characteristics of the virtual environment, how it responds to people's actions, and very importantly their own self-representation within the environment. In other words, it is a valid scientific problem to explore the relationship between the subjective and behavioural phenomenon that we call 'presence' and the degree of 'system immersion'. Some particular examples that our group have explored are:

- The influence of having a virtual body (an avatar) on presence (Slater & Usoh, 1993);
- The influence of having dynamic shadows on presence (compared to not having shadows) (Slater, Usoh & Chrysanthou, 1995);
- The influence of one particular method of 'walking' in an environment compared to others (Slater, Usoh & Steed, 1995);

Measuring Presence

- The extent to which the environment (or interaction paradigms) that require semantically appropriate body movement influences presence (Slater, Steed, McCarthy & Maringelli, 1998).

We have also studied personal factors (Slater, Usoh & Steed, 1994) - such as the extent to which a person's field dominance (e.g., visual, auditory) influences their sense of presence, however, this is not part of system immersion, and so will not be considered further here.

People working in the area of presence are trying to map out an equation. On the left-hand side is the presence response. On the right hand side are the components of system immersion. Barfield and Hendrix' study of frame rate is a further example (Barfield & Hendrix, 1995), and the Welch study of pictorial realism and interactivity is another (Welch, Blackman, Liu, Mellers & Stark, 1996). On the right hand side of the equation are measurable characteristics of a system. On the left hand side is some attempt at measurement of this phenomena that we call 'presence'.

The endeavour to map out this equation has very practical consequences - if we knew the equation then we could build systems that 'maximised' presence (if anyone cared about this). Of course, the equation might be application dependent. Of course there will be differences in the equation coefficients and constants between one person and another, due to their different psychological make-up. But the goal is clearly defined, even if we don't exactly know how to do this (because we don't have a good way to measure presence itself). The research is worth-while for another reason - the answers are not obvious. More realism across the board might not equal more presence - different variables will have different marginal effects. We will no doubt find, as Ellis requires (Ellis, 1996), an equation that allows us to trade off factors against each other, while still maintaining the same level of presence. This again is important for system engineering - since one can envisage manipulating a function under constraints (of cost) - for example, sacrificing one aspect of system immersion, to increase another, while keeping presence constant, and reducing cost.

Measuring Presence

Now the approach in the W&S paper is in fact is very different from this. They define presence as quoted above. The paper then considers a number of factors thought to influence presence. These are:

Control factors. Degree of control, immediacy of control, anticipation of events, mode of control, physical environment modifiability;

Sensory factors. Sensory modality, environment richness, multimodal presentation, consistency of multimodal information, degree of movement perception, active search;

Distraction factors. Isolation, selective attention, interface awareness;

Realism factors. Scene realism, information consistent with objective world, meaningfulness of experience, separation anxiety/disorientation.

There is not space to go into the detailed meaning of all of these, and readers are referred to the original paper. An important point about these factors is that they are all *subjectively defined*.

For example, consider the following questions relating to *control factors* (the numbering is not the original numbering in the W&S questionnaire):

1. How much were you able to control events?
2. How responsive was the environment to actions that you initiated (or performed)?
3. How well could you move or manipulate objects in the virtual environment?

Here are some questions in the *sensory factors* category:

4. How much did the visual aspects of the environment involve you?
5. How compelling was your sense of moving around the virtual environment?
6. How well could you examine objects from multiple viewpoints?

Examples of the *distraction factors* category are:

Measuring Presence

7. How aware were you of events occurring in the real world around you?
8. How aware were you of your display and control devices?
9. How distracting was the control mechanism?

Finally some examples of the *realism factors*:

10. How inconsistent or disconnected was the information coming from your various senses?
11. To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?
12. How much did your experiences in the virtual environment seem consistent with your real-world experiences? (Also a control factor question).

Now of course these questions are fine in themselves. But note that throughout they elicit the *opinions* of the experimental subjects about these matters. Let's consider an example to demonstrate why this may not be useful. Suppose there is a CAVE-style virtual environment, with a visual display on all six walls, and the freedom of people to walk around the CAVE with full collision detection, etc. The subjects hold a tennis racquet which is tracked. Virtual tennis balls fly towards subjects, and they have to hit as many as possible. Subjects individually enter the environment, and carry out the task with varying degrees of success. For two such subjects one reports a high degree of ability to control events, and another reports a very low ability. Of course the system is the same in both cases. What has determined the difference in response is nothing at all to do with the immersive system, but is due to differences in the individuals, their experience, psychological make-up, dexterity, and so on. They each report a different immersive response (in W&S's meaning). Note that there is nothing here to suggest that one would have a higher or lower sense of presence than the other. The one who did not perform so well might well have had a lower self-expectation of performance, and the fact this person was 'not able to control events' may have enhanced their sense of presence because they experience the same in real life. Now this would be caught by another question (12. above). But on this question the two people are likely to give the same answer - the more dexterous person and the less skillful person would each find the virtual world consistent with real life.

Measuring Presence

Consider question 7. One person might answer that they were very much aware of events occurring in the real world, and the other report a very low level of awareness. Again, this has nothing to do with the system - since both have the same chance to become aware. The different answers are to do with the individuals' personal responses, due to their different ways of making sense of the world, due to their different perceptual and psychological make-ups.

W&S say 'We believe that the strength of presence experienced in a VE varies both as a function of individual differences and the characteristics of the VE'. But the presence questionnaire (PQ) by itself makes it impossible to separate out these two influences. We get the immersive tendencies questionnaire (ITQ) (which provides a measure of individual differences) and the presence questionnaire - but which also elicits people's subjective responses to *various aspects of the system*. We would like to be able to independently vary these two factors: examine the variation of the presence response amongst people of similar 'personality type' across different system factors (e.g., different frame rates), and alternatively examine the variation in responses of different types of people to the same system configuration. Measures of system immersion, and independent measures of individual personal characteristics and experience should be sufficient for this purpose. The presence questionnaire PQ confounds them.

There are 32 questions of the type above in the PQ. Each question is answered on a 1 to 7 scale. The strategy is to ask questions that are believed (from the literature or from other experimental research) to influence presence. The overall score is the sum of the answers to all 32 questions. Note that none of the questions is directly about presence. This questionnaire therefore cannot be used to study the factors, according to W&S, that influence presence. The presence score is constructed out of those factors. It is their sum. The operational definition of presence employed by W&S (notwithstanding its conceptual definition) is that it is a completely deterministic function of its influencing factors. We cannot separate out two different types of entity: a measure of presence, and independently a measure of factors that might influence it. Changes in the latter automatically cause changes in the measured response, because that is how the measured response is constructed.

Measuring Presence

Putting this another way, suppose that x_1, x_2, \dots, x_n are factors thought to be influencing presence p .

Then according to the W&S model:

$$p = x_1 + x_2 + \dots + x_n \quad (\text{EQ 1})$$

If I then ask the question, ‘what is the relationship between presence and factor 1’ - there’s no experiment to do, the answer is given by (EQ 1).

W&S point out the correlations between the individual factors (for example, question 1) and the overall score across some experiments. For example, the correlation coefficient with question 1 is 0.43 in a sample of 152, which is reported as significant at $p < 0.001$. Suppose x_1, x_2, \dots, x_n are n independent random variables with means μ_i and variances σ_i^2 , ($i = 1, \dots, n$). Let

$$p = \sum_{i=1}^n x_i \quad (\text{EQ 2})$$

It is easy to show that the covariance between, say, x_i and p is σ_i^2 . If we let ρ_i be the correlation coefficient between x_i and p then it follows that:

$$\rho_i^2 = \frac{\sigma_i^2}{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2} \quad (\text{EQ 3})$$

The meaning of this is that the correlations between the individual questions and the overall sum are determined solely by the variances of answers to the individual questions. A question that has high variability across subjects in the experiment will obviously correlate more highly with the sum than one with a lower variance. This gives no information at all about the influence of these variables on ‘presence’ - only about their relative variation amongst each other for this particular set of experimental subjects. For example, if a person’s age was one of the factors that went into the sum,

Measuring Presence

and an experiment was conducted where age had the highest variance, then it would be concluded that age is a significant contributory factor to the presence construct.

It could be argued that all questions are rated on a 1 to 7 scale, so that the variances might be similar. Let's make the further simplifying assumption that all the variances are equal. In this case (EQ 3) reduces to:

$$\rho_i = \frac{1}{\sqrt{n}} \quad (\text{EQ 4})$$

If the x_i are the answers to the 32 questions, then $n=32$, and $\rho_i = 0.18$.

It is clear that just by construction of the scale there is bound to be correlation between the individual questions and their sum. Moreover, if we use Fisher's z-statistic for the correlation coefficient (for example, Dunn & Clark, 1974, p244) we find that the probability of the observed correlation coefficient being in the range 0.02 to 0.33 is approximately 0.95. W&S quote significance levels (for a null hypothesis of *zero correlation* between an individual question and the overall sum) of 0.001 for values of the observed correlation as low as 0.33. A zero correlation null hypothesis is not appropriate in these circumstances.

Now the real situation is worse than this. Of course, the factors will not be independent - they will be correlated amongst themselves. In this case (EQ 3) will not hold, but many of the questions will have a higher theoretical correlation with their sum. This is to do with the structure of the construct, and nothing to do with any empirical relationships around the issue of presence.

Overall then, notwithstanding the reliability and consistency of the questionnaire across several experiments, I cannot use it in my research. It does not give a measure of presence that is constructed independently from the factors that might influence it. It does not, from my point of view, measure presence at all - rather it is a measure of a people's responses to various aspects of a system, which of course is likely to be correlated with other measures of presence. I would rather

Measuring Presence

attempt (even if not be successful) to find a measure of presence based on a methodology that is directly concerned with the concept in itself, as usefully summarised in the earlier quote from W&S.

The argument of W&S that their measure is valid is open to doubt. For example, they report some experiments relating task performance to their construct. First, there is no logical requirement for there to be a positive relationship between presence and task performance. Task performance depends on 'user interface' and personal skills and experience - an individual can be in the real world doing a task (e.g., trying to draw money from an ATM) be totally present in any reasonable definition of the word, yet exhibit poor task performance either because of the interface to the machine, or for some other personal reason. But even accepting that presence and task performance should be correlated, W&S report that out of 4 experiments, their presence construct was positively and significantly related to task performance in two of them, but not in the other two. In another experiment on natural modes of interaction, again their score did not result in the expected significant differences between two experimental groups. Even on the immersive tendencies questionnaire, 'only two of the four experiments resulted in a significant correlation between ITQ and PQ scores'. The four experiments were combined together and a significant overall correlation was found. Their conclusion that 'presence, as measured by the PQ, is a valid construct' is not supported by the evidence that the authors present.

The reasons have been set out why I would not use the W&S presence questionnaire. I would use the ITQ, since this stands alone as an attempt to measure important psychological characteristics of individuals. I would rather not use questionnaires at all. In many ways very profound insights into the nature of presence can be gained from ethnographic studies (for example, McGreevy, 1993). However, at the end of the day I use questionnaires because, for the time being, I do not know what else to do, and in order to construct predictive equations, concerned with how presence varies with other factors in groups of people, some method of quantification is necessary.

References

Barfield, W. and Hendrix, C. (1995) The Effect of Update Rate on the Sense of Presence within Virtual Environments, *Virtual Reality: The Journal of the Virtual Reality Society*, 1(1) 3-16.

Dunn, O.J and Clark, V.A. (1974) *Applied Statistics: Analysis of Variance and Regression*, John Wiley and Sons, New York.

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*, MIT Press, 5(2), 247-259.

Held, R.M. and N.I. Durlach (1992) Telepresence, *Presence: Teleoperators and Virtual Environments*, 1, winter 1992, MIT Press, 109-112.

McGreevy, M.W. (1993) The Presence of Field Geologists in a Mars-Like Terrain, *Presence: Teleoperators and Virtual Environments*, 1(4) fall 1993, MIT Press, 375-403.

Sheridan, T.B. (1992) Musings on Telepresence and Virtual Presence, *Telepresence, Presence: Teleoperators and Virtual Environments*, 1, winter 1992, MIT Press, 120-126.

Slater, M. and Usoh, M. (1993) Presence in Immersive Virtual Environments, *IEEE Virtual Reality Annual International Symposium (VRAIS)*, September 18-22, Seattle, Washington, 90-96.

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

Slater, M., Usoh, M., Chrysanthou, Y. (1995) The Influence of Dynamic Shadows on Presence in Immersive Virtual Environments, *Second Eurographics Workshop on Virtual Reality*, M. Goebel, ed., Monte Carlo, Jan., 1995.

Slater, M., Usoh, M., Steed, A. (1995) Taking Steps: The Influence of a Walking Metaphor on Presence in Virtual Reality, *ACM Transactions on Computer-Human Interaction (TOCHI)* 2(3) September, 201-219.

Slater, M., Wilbur S. (1997) A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments, *Presence: Teleoperators and Virtual Environments*, 6(6) 603-616, MIT Press.

Slater, M., Steed, A., McCarthy, J. and Marinelli, F. (1998) The Influence of Body Movement on Presence in Virtual Environments, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3), 469-477.

Welch, R.B., Blackman, T.T., Liu, A., Mellers, B.A., Stark, L.W. (1996) The Effects of Pictorial Realism, Delay of Visual Feedback, and Observer Interactivity on the Subjective Sense of Presence, *Presence: Teleoperators and Virtual Environments*, MIT Press, 5(3), 263-273.

Measuring Presence

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