

Jonathan Freeman
jfreem@essex.ac.uk

S. E. Avons
Department of Psychology
University of Essex
Wivenhoe Park
Colchester, Essex
CO4 3SQ, UK

Don E. Pearson
Department of Electronic Systems
Engineering
University of Essex
Wivenhoe Park
Colchester, Essex
CO4 3SQ, UK

Wijnand A. IJsselstein
IPO
Center for Research on User-
System Interaction
Eindhoven University of Technology
Den Dolech 2
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence

Abstract

We report three experiments using a new form of direct subjective presence evaluation that was developed from the method of continuous assessment used to assess television picture quality. Observers were required to provide a continuous rating of their sense of presence using a handheld slider. The first experiment investigated the effects of manipulating stereoscopic and motion parallax cues within video sequences presented on a 20 in. stereoscopic CRT display. The results showed that the presentation of both stereoscopic and motion-parallax cues was associated with higher presence ratings. One possible interpretation of Experiment 1 is that CRT displays that contain the spatial cues of stereoscopic disparity and motion parallax are more interesting or engaging. To test this, observers in Experiment 2 rated the same stimuli first for interest and then for presence. The results showed that variations in interest did not predict the presence ratings obtained in Experiment 1. However, the subsequent ratings of presence differed significantly from those obtained in Experiment 1, suggesting that prior experience with interest ratings affected subsequent judgments of presence. To test this, Experiment 3 investigated the effects of prior experience on presence ratings. Three groups of observers rated a training sequence for interest, presence, and 3-Dness before rating the same stimuli as used for Experiments 1 and 2 for presence. The results demonstrated that prior ratings sensitize observers to different features of a display resulting in different presence ratings. The implications of these results for presence evaluation are discussed, and a combination of more-refined subjective measures and a battery of objective measures is recommended.

1 Introduction

In any communication medium, a message is sent from a source to a receiver. In many cases the message describes events taking place in an environment remote from the receiver; this is true of all fictional literature and most televisual services. The receiver, or observer, is physically present in one environment but is engaged by—and responds to—the remote environment. The observer's subjective sensation of "being there" in the remote environment is termed *presence*. As presence increases, the observer becomes more aware of and engaged by the mediated environment, and less aware of the environment in which he or she is physically located. The development of visual media such as photography, cinema, and television can be viewed as attempts to increase realism by increasing the size and fidelity of the displays. In the case of interactive displays, the observer can actively explore the mediated environment (for

example, selecting viewpoints by turning the head). Thus, the perceptual linkage between the observer and the displayed environment is enhanced. The main concern of our research program is to investigate the extent to which the fidelity of the visual display (particularly stereoscopic information) determines the sensation of presence in the viewer. There are potentially many ways of addressing this question. In the present paper, we evaluate the use of continuous subjective ratings of presence.

Recent accounts of presence propose that it is a complex, multidimensional construct (Barfield, Zeltzer, Sheridan, & Slater, 1995), a subjective sensation much like mental workload (Sheridan, 1992). Several papers have discussed the determinants of presence, broadly agreeing on the concepts if not the terminology. The major analyses have come from Sheridan (1992), Ellis (1996), Slater and Usoh (1994), Heeter (1992), and Welch, Blackmon, Liu, Mellers, and Stark (1996). Our investigation of presence is best framed in Sheridan's analysis.

Sheridan (1992) identifies presence as having three major determinants:

- (i) the extent of sensory information (i.e., the amount of useful and salient sensory information concerning a cue available to the appropriate sensors of the observer)
- (ii) the control of the relation between sensors and the display (i.e., the degree of control a participant has over the positioning of his or her sensors within the environment)
- (iii) the ability to modify the physical environment (i.e., the degree to which a participant is free to modify objects and their positions within the environment)

Other researchers have combined (ii) and (iii) into one determinant, "user interaction" (Zeltzer, 1992). The experiments presented in this paper test Sheridan's prediction that adding further ecologically important sensory information to a display, in the form of stereoscopic and motion-parallax cues, enhances presence.

1.1 Measuring Presence

A number of studies on presence evaluation have utilized direct subjective presence assessment. Slater and Usoh (1994) assessed the sense of presence experienced by participants who were instructed to perform various tasks in six rooms off a corridor in a virtual environment. They assessed presence in a number of ways but reported only participants' responses to a questionnaire. The question asked was, "To what extent did you experience a sense of being 'really there' inside the virtual environment?" Participants could respond along a six-point scale ranging from "not at all really there" to "totally there."

Similarly, Slater, Usoh, and Steed (1994) conducted experiments to assess the influence of both "internal" and "external" factors on the reported level of presence in an interactive, immersive virtual environment, based on the fairy tale *Beauty and the Beast*. Slater et al refer to "external factors" as the parameters of a virtual environment (e.g., its field of view or "degree of interactivity"), whilst they describe internal factors as affecting individuals' responses and perceptions to identical external stimuli. The measure of presence was based on responses to three questions using a post-experimental questionnaire. Participants were asked to rate

- (i) their *sense of being there* (in the computer-generated world),
- (ii) the extent to which there were times during the experiment when the computer-generated world became "reality" for the participants (whereby they almost forgot about the "real world" outside), and
- (iii) whether they thought of the computer-generated world as *something they had seen* or *somewhere they had visited*.

Ratings were on Likert scales (1 to 7). A presence score for each participant was a simple count of the number of scores of 6 or 7 in response to the three questions and so had a range of 0 to 3. Results from these studies demonstrated that there is some association between a participant's dominant representation style (internal factors) and their reported sense of presence.

Hendrix and Barfield (1996a) researched the effect of display parameters on presence ratings within virtual environments. Presence was evaluated using a questionnaire that contained specific questions addressing each of the following manipulations:

- (i) the presence or absence of head tracking
- (ii) the presence or absence of stereopsis
- (iii) geometric field of view (GFOV)

Each participant in an experiment experienced two virtual worlds and was able to return to either as many times as they wanted to answer the questionnaires. The presence questions asked for each manipulation were

- (i) "If your level of presence in the real world is '100' and your level of presence is '1' if you have no presence, rate your level of presence in this virtual world."
- (ii) "On a scale of 1 to 5, how strong was your sense of presence, 'being there,' in the virtual environment? (where 1 = very much so, and 5 = not at all)."

Where stereopsis was an independent variable, questions were also asked regarding the visual realism of the environment, the realism of the displayed depth, and whether observers felt that they could have reached into the environment to grasp an object. If head tracking was manipulated, questions were also asked regarding how realistically the environment interacted with the observers, and how realistically the world moved in response to the observers' head movements. If GFOV was manipulated, questions were also asked regarding the realism of the environment, whether the environment appeared to be compressed or magnified, whether the world seemed too narrow or wide as compared with the real world, and whether objects appeared proportionally correct in terms of size and distance in relation to the observers and other virtual objects.

Hendrix and Barfield's (1996a) results showed that the reported levels of presence were significantly increased when head-tracking and stereoscopic cues were added and when wider fields of view (50 deg. and 90 deg.) were presented as opposed to a narrow field of

view (10 deg.). Hendrix and Barfield (1996b) used the same method to evaluate the participants' perceptions of presence within auditory virtual environments, finding that increasing the display realism using spatialized sound increased the reported perceptions of presence. Hendrix and Barfield reported that responses to the questions used in the above studies were relatively consistent and concluded that direct subjective evaluation of presence is an adequate means of assessment. It is, however, important to note the dependence of the above studies on a simple rating scale of presence, and that potential problems exist in the observers' understanding of the term and in defining rating scale endpoints.

An important distinction between the work of Slater and colleagues (Slater & Usoh, 1994; Slater et al. 1994) and that of Hendrix and Barfield (1996a) is that Slater's studies used independent groups designs while Hendrix and Barfield's used repeated measures designs. While an independent-groups design might minimize the chance of procedural bias affecting results, it is also the case that a repeated-measures design increases the sensitivity of a measure. This difference is potentially important, and we return to the topic in the general discussion.

1.2 Continuous Presence Assessment

A potential problem with post-test evaluation is that of inaccurate recall and memory effects such as recency. Aldridge, Davidoff, Ghanbari, Hands, and Pearson (1995) demonstrated a recency effect in observers' post-test overall ratings of TV picture quality. The use of a continuous assessment methodology has the benefit that data is collected online, thus minimizing the contribution of any memory effects.

We have adapted the method of continuous evaluation of TV picture quality (developed under the EC RACE-MOSAIC project (ITU-R, BT 500-7) to presence assessment (IJsselsteijn, Freeman, Avons, Davidoff, de Ridder, & Hamberg, 1997; IJsselsteijn, de Ridder, Hamberg, Bouwhuis, and Freeman, 1998; Freeman, Avons, & Davidoff, 1997a; Freeman, Avons, Davidoff, & Pearson, 1997b). The continuous assessment methodology has been shown to be reliable and consistent in

a number of studies into digitally coded, TV-picture-quality evaluation (Hamberg & de Ridder, 1995; de Ridder & Hamberg, 1997; Aldridge et al., 1995). Our adaptation of this method requires that observers provide a continuous rating of their sense of presence using a handheld slider while being presented with the displayed stimulus. Observers are instructed to move the slider up when they feel an increase in presence and to move it down when they feel a decrease in presence. *Presence* is defined for observers as “a sense of being there” in a displayed scene or environment. No rating scale endpoints are provided for observers, although informal reports from observers in previous experiments indicate that they treat the top of the scale as meaning “completely there” and the bottom of the scale as “not at all there” in the displayed scenes. In addition, to aid observers in providing their ratings, a training period of three minutes is given. In this period, observers are asked to provide presence ratings for stimuli variations similar to those they experience in the experiment for which they are being trained. This enables observers both to get accustomed to the rating device and to experience the range of presence they are likely to experience. The handheld slider is a 10 cm long analog scale with a small slider positioned in a central groove running the length of the scale. Observers do not look at the scale and rely on haptic feedback to indicate slider position. Observers are shown how to hold the slider (with a finger at the top of the scale and thumb at the base) such that the haptic feedback they receive while providing ratings is optimized. The analog scale gives an 8-bit digital readout, ranging from a minimum of 0 to a maximum of 255. (These scale scores are the units in which we report the results of the experiments presented in this paper.) The temporal aspects of observers’ ratings are not reported in this paper but are discussed in IJsselstein et al. (1998).

A major consideration for any psychophysical measurement technique is that it does not influence the sensation being evaluated. In the case of presence, the observer is asked to rate the extent to which he or she feels present in the virtual environment. Concurrent ratings could interfere in two ways: by imposing an additional mental load, thus diverting attention away from the dis-

play, or by forcing the observer’s attention back to the real-world environment (e.g., by requiring the operation of some external device). The rating procedure used here was simple and undemanding to use, and the task was well practiced in all observers before measures were taken. In addition, the observers used a handheld rating device that was in their egocentric space, but not fixed in the laboratory environment, and that they were not required to look at while making ratings. Thus, we attempted to minimize the influence of the measurement technique. The extent to which the concurrent rating technique influences presence ratings is an interesting question, but the present experiments require only relative—and not absolute—measures of presence, since the rating procedure was the same in all conditions.

2 Experiment 1—Effects of Viewing Condition and Motion Parallax on Presence Ratings

This experiment was designed to examine the effects on subjective presence ratings of manipulations of stereoscopic and motion-parallax cues to depth within stereoscopically presented video sequences.

Previous experiments have shown that observers watching a stereoscopic video will provide time-varying presence ratings that are correlated with variations in both the rated depth of the stimulus and the rated naturalness of that depth. Consistent temporal variation was observed for the same sequences in two laboratories, demonstrating the reliability of the method if the instructions and procedure are kept constant (Freeman et al., 1997a; IJsselstein et al., 1998). In support of theories of the determinants of presence (e.g., Sheridan, 1996), results of this earlier work suggested that observers produced higher ratings of presence in sections of the stimulus film that were stereoscopically presented and in scenes with camera motion (and, hence, motion parallax). However, the early experiments were limited, as the presentation of stereoscopic and monoscopic sections of the film was not randomly varied, and there were no controls on the amount of motion in the scenes.

2.1 Method

2.1.2 Stimuli In this experiment, the viewing condition (stereoscopic/monoscopic) was fully controlled and counterbalanced, and the amount of motion and the scene content in each sequence was controlled as well as possible, given that there was a limited amount of source material. Three 30 sec. sections of film were selected according to the amount of motion they contained:

- (i) observer motion, in which the camera moved laterally and turned to keep the subject in view, without zooming
- (ii) scene motion, in which the actors moved within the scene, but the camera was still
- (iii) minimal motion, in which the maximum motion within the sequence was a small gesture by an actor or some leaves gently blowing in the breeze, again with no camera motion

The distance from the camera to the actors and objects in each of the sequences was closely matched, and was in all cases in the range of 2 to 10 meters. (The range utilized did not differ significantly between sequences.) For stereoscopic presentation of stimuli, two streams of synchronized video were played for the observers, one stream of stimuli, two streams of synchronized video were played for the observers, one stream for the left eye and one for the right eye. For monoscopic presentation, the left-eye video stream was presented to both eyes.

2.1.3 Observers Twelve university students (six men, six women, average age of 22 years) volunteered to participate in the experiment. All had normal or corrected-to-normal vision. All observers had a stereoacuity of 30 sec-arc or better, as tested on the RANDOT® random dot stereotest. (This test was used to establish the observers' stereoacuties in all the experiments reported here.)

2.1.4 Apparatus Observers viewed the stimulus films on an AEA Technology 20 in. stereoscopic display consisting of two BARCO CPM 2053 color monitors (50 Hz PAL) with polarized filters in front of each. (See

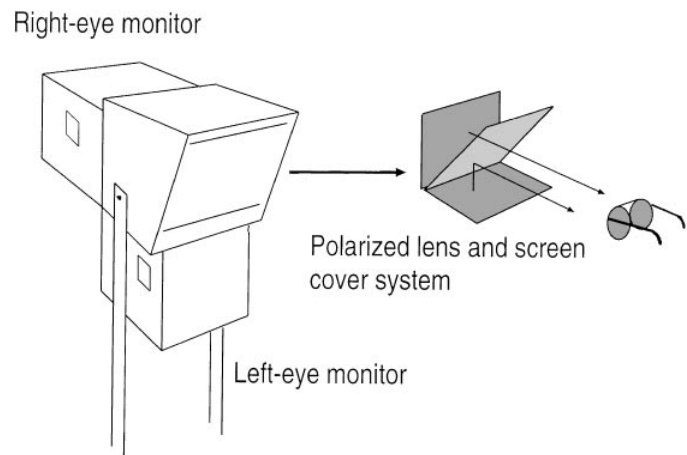


Figure 1. Schematic diagram of the AEA Technology 20 in. stereoscopic display. Right- and left-eye views are presented at the same time (i.e., time parallel), and polarization is used to separate the views.

figure 1.) Observers viewed the display wearing polarized spectacles. The handheld slider used by observers to provide their ratings was connected to a standard personal computer (PC), running software that sampled and stored the handheld slider position at a rate of 5 Hz. Two synchronized Panasonic M2 (A750-B) video players provided the video input for the display.

2.1.5 Procedure One observer at a time took part in the experiment. On arrival at the laboratory, observers were seated 80 cm (approximately two picture heights) away from the stereoscopic display. They were then asked to read written instructions that explained their task and that defined presence as a sense of being there in a displayed scene. The instructions were reiterated verbally by the experimenter. Observers were asked to equate the scale on the handheld slider with that presented on the instruction sheet. The lights in the room were switched off after observers had communicated a thorough understanding of their task. All observers first viewed a 3 min. practice sequence of stereoscopic film containing both stereoscopic and monoscopic scenes and variable amounts of motion—including a view of a rally track from the rear seat of a racing car and studio shots of a presenter, some with motion and others practically still—while providing presence ratings. Data from the practice sessions was not collected.

Table 1. *Composition of Stimuli Video Tapes for All Experiments*

Left Eye Right Eye	Left Eye 1 Left Eye 1	Left Eye 2 Left Eye 2	Left Eye 3 Left Eye 3	Left Eye 1 Right Eye 1	Left Eye 2 Right Eye 2	Left Eye 3 Right Eye 3
	monoscopic			stereoscopic		
	30 seconds	30 seconds	30 seconds	30 seconds	30 seconds	30 seconds
	180 secs					

The design of this experiment was fully repeated measures, such that all observers saw all the motion sequences both monoscopically (both eyes receiving the left-eye view) and stereoscopically (each eye receiving its appropriate view). The order of the motion sequences was fully counterbalanced across observers. Half the observers saw the sequences first stereoscopically, then monoscopically, and half in the reverse order. The composition of a test sequence for an observer in the mono-first condition is shown in Table 1, labeling the levels of motion as 1 (observer motion), 2 (scene motion), and 3 (minimal motion). Each of the twelve observers saw the film sequences in a different counterbalanced order.

2.2 Results

Previous experiments have shown that observers typically take approximately ten seconds from a scene change to reach a steady presence rating; thus, for the purpose of analysis, the first ten seconds of ratings for each motion sequence were ignored. Means for each motion sequence per observer were then calculated, and these provided the input data for statistical analysis.

Figure 2 shows the group mean presence ratings for each motion sequence split by viewing condition. It is clear from the graph that both viewing condition and motion parallax affected observers' presence ratings. A two-factor (2×3) repeated-measures ANOVA was run and showed a main effect of viewing condition ($F(1, 11) = 17.970, p < 0.001$) and of motion parallax ($F(2, 22) = 3.883, p < 0.05$). Presence ratings were higher for stereoscopically presented sequences and for scenes containing camera motion. A post-hoc analysis (Tukey-HSD) of the motion-parallax effect revealed that only

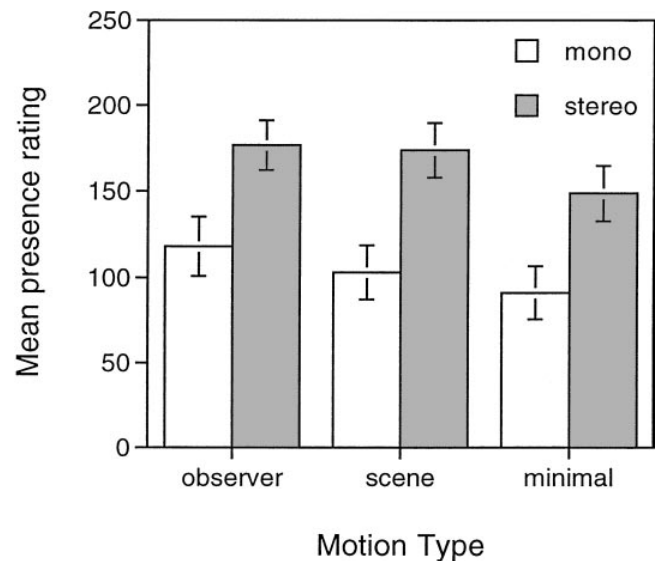


Figure 2. Data from Experiment 1. Effects of viewing condition and motion parallax on group mean subjective ratings of presence. Bars indicate standard errors.

the difference between the observer motion and minimal motion conditions was significant. There was no significant interaction between the viewing condition and motion-parallax effects ($F(2, 22) = 0.700, n.s.$).

2.3 Discussion The results of Experiment 1 confirm previous work showing that subjective feelings of presence are enhanced by stereoscopic stimulus presentation (Hendrix & Barfield, 1996a; Freeman et al., 1997a, 1997b; IJsselstein et al., 1998). In addition, the results are supportive of theories of the determinants of presence in that observers provided higher presence ratings when more sensory information was available.

The absence of an interaction between the two factors suggest that stereoscopic presentation and motion parallax operate independently on observers' sense of presence. This is an informative result given that stereopsis and motion parallax supply essentially the same information about the depth content of a scene (Rogers & Graham, 1982). If information about the depth structure of a displayed scene was the sole requirement for a sense of presence, we would expect that this redundancy of the stereoscopic and motion-parallax cues would lead to a statistical interaction. The absence of this interaction can be explained in three ways. First, it is possible that, in addition to providing depth information, one or both of the cues was fulfilling another role (e.g., stereoscopic information effectively disguises the picture plane). Second, it is possible that, because the effectiveness of stereopsis is optimal for near objects and because motion parallax has a less limited range of effectiveness, the two were effectively working in different parts of the image. The third explanation derives from the fact that in natural vision the two cues occur together and, hence, that the naturalness of the image was increased when both cues were present. The high correlation between naturalness of presented depth and presence found in earlier experiments supports the third explanation (Freeman, Avons, & Davidoff, 1997a; IJsselstein, de Ridder, Hamberg, Bouwhuis, & Freeman, 1998).

The magnitude of the difference in ratings among the viewing conditions and the various levels of the motion cue would suggest that stereopsis was a more important determinant of presence than motion. However, with the available stimuli, the presence or absence of stereoscopic information was better controlled than the degree of motion within the stimuli. The experimenter had full control over the viewing conditions for the stimuli, but had to select the most-controlled possible motion sequences from the available stereoscopic video. To explore this question more fully, sequences with controlled motion-parallax variations must be devised.

A number of concerns relate to the experimental procedure. Observer recruitment relied on an advertisement for a "3DTV experiment," all observers underwent a stereoacuity test prior to inclusion in the experiment, and observers were required to wear polarized spectacles

while viewing the display. Hence, observers were led to believe that stereoscopic presentation would occur. This may have biased them to consider stereoscopic viewing to be an important contribution to presence. In the absence of a firm understanding of the construct, there was potential for observers to assume that they would not feel "present" in the absence of three-dimensional depth within the stimuli.

To minimize the potential for any priming of expectation, Experiments 2 and 3 were advertised as "new media research", and observers were tested for stereoacuity *after* participating in the experiment.

3 Experiment 2—Effects of Viewing Condition and Motion Parallax on Interest Ratings and Presence Ratings

In Experiment 1, motion parallax was varied by selecting different scenes with different degrees of motion. It is possible that other characteristics of the scenes, such as their interest, influenced the ratings found. This experiment was designed to investigate whether the effect of motion parallax on presence found in Experiment 1 could be attributed in part to variations in interest. Experiment 2 consisted of two parts. In the first part (referred to as Experiment 2a), observers rated stimuli for interest. Following this, as a check on the consistency of the results of Experiment 1, the same observers rated the same stimuli for presence in Experiment 2b.

3.1 Experiment 2a—Interest Ratings

The same stimuli as used for Experiment 1 were presented to a new group of observers who were required to continuously rate how interesting the sequences were. Interest was chosen as the dependent variable because the term is broad enough to tap into a range of potential factors (e.g., novelty, involvement, and engagement by the narrative content of the sequences) that are likely to affect the observers' sense of presence. This point is important because asking for interest ratings did not limit observers to rating the presented sequences solely on the quality of their narrative

structure, which was vastly reduced by extracting short sequences and presenting them without accompanying audio. Rather observers were free to provide a more general rating of interest in the whole presentation, which could encompass the narrative structure and scene content as well as the physical appearance of the display (3-D versus 2-D).

3.1.1 Method Observers. Twelve university students (six men, six women, average age of 25 years) volunteered to participate in the experiment. All had normal or corrected-to-normal vision and a stereoacuity of 30 sec-arc or better.

3.1.1.1 Apparatus and Materials. These remained identical to those in the previous experiment.

3.1.1.2 Procedure. The procedure was identical to the previous experiment, except for the instructions observers received: in this experiment observers were requested to provide continuous ratings of “how interesting they found the film.” The same practice sequence as in the preceding experiment was again utilized. The design of this experiment was identical to that of the preceding experiment.

3.1.2 Results and Discussion Figure 3 shows the group mean ratings of interest for each motion sequence split by viewing condition. Somewhat higher interest ratings were obtained when the sequences were presented stereoscopically, but there was little effect of motion parallax.

A two-factor (2×3) repeated-measures ANOVA was run, with the input data being each observer's average rating across a motion sequence (excluding the first ten seconds as for the previous experiment). This analysis showed that the effect of viewing condition illustrated in the graph below was not statistically significant ($F(1, 11) = 2.021, p = 0.183$). Neither was there an effect of motion parallax ($F(2, 22) = 0.334, n.s.$). In addition, there was no significant interaction between the two factors ($F(2, 22) = 0.866, n.s.$).

Stereoscopic presentation of stimuli showed only a small, nonsignificant increase in interest ratings com-

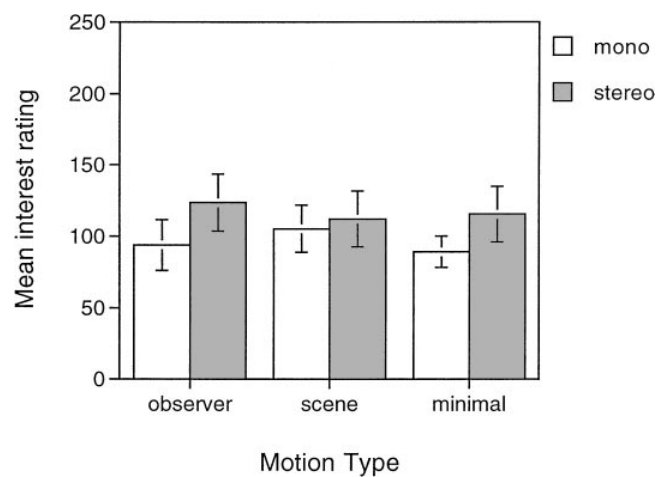


Figure 3. Data from Experiment 2a. Effects of viewing condition and motion parallax on group mean subjective ratings of interest. Bars indicate standard errors.

pared with monoscopic presentation. There was also no effect of motion parallax. Although we are not reporting on the temporal aspects of the ratings, the temporal variation in the interest ratings from this experiment was different to the temporal variation found in the presence ratings in Experiment 1. These results on interest ratings, therefore, differ from those found with presence ratings in Experiment 1, and we conclude that rating interest is not equivalent to rating presence.

3.2 Experiment 2b—Presence Ratings

The second part of this experiment was a replication of Experiment 1. The twelve observers who provided interest ratings in Experiment 2a were requested to view the stimuli again, this time providing continuous ratings of presence. The same written instructions as used in Experiment 1 were given to the observers, and these instructions were again reiterated verbally by the experimenter. The differences between the procedures experienced by observers in Experiment 2b and in Experiment 1 were that observers in Experiment 2b

- had already seen the stimuli and provided interest ratings for them,
- had viewed 180 sec. more stereoscopic TV,

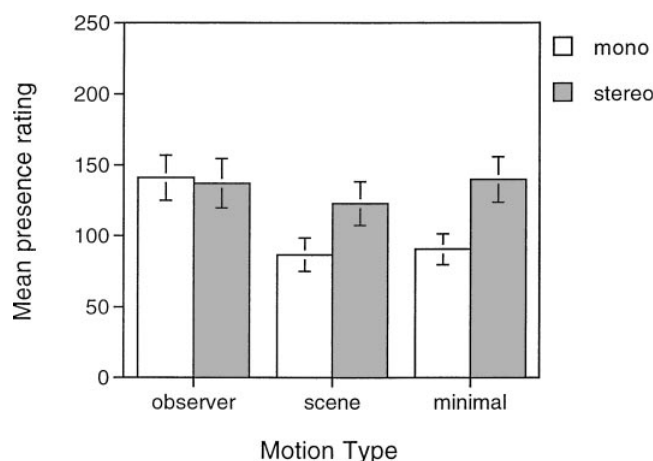


Figure 4. Data from Experiment 2b. Effects of viewing condition and motion parallax on group mean subjective ratings of presence. Bars indicate standard errors.

- (c) had received 3 min. practice in rating interest rather than presence,
- (d) were recruited through advertisements for “new media research” rather than “3DTV experiment”, and
- (e) were tested for stereoacuity after completing the experiment.

3.3 Results and Discussion

Figure 4 shows the group mean ratings of presence for each motion sequence split by viewing condition. A comparison of Figure 4 with Figure 2 reveals clear differences. The most obvious difference is the reversal of the stereoscopic advantage in the observer motion sequence shown in Figure 2 to a monoscopic advantage in ratings by observers who had previously rated the same sequence for interest (Figure 4).

This difference was confirmed statistically in a two-factor (2×3) repeated-measures ANOVA. Overall, there was an advantage for stereoscopic viewing, but this failed to reach significance ($F(1, 11) = 2.365$, $p = 0.152$). As for Experiment 1, there was a significant effect of motion ($F(2, 22) = 7.367$, $p < 0.01$), but here also was a significant interaction between the stereo and motion effects ($F(2, 22) = 4.031$, $p < 0.05$).

There are three possible explanations of the difference in the presence ratings obtained for identical stimuli in Experiment 1 and Experiment 2b. One explanation is that the methodology is simply unstable. This is unlikely given that similar ratings of presence were obtained in two independent laboratories in different countries in an earlier experiment using the same methodology (Freeman, Avons, & Davidoff, 1997a; IJsselstein et al., 1998).

A second possible explanation is that presence ratings vary as a function of the number of exposures to the same stimulus. For example, the presence advantage for stereoscopic stimuli may become attenuated over repeated presentations. This explanation is not formally tested in the present study, but it cannot account for all our results. In addition, it should be noted that repeating scenes does not affect presence ratings, since, in all the present experiments, the same sequences were shown monoscopically and stereoscopically, and no effects of order of presentation were found.

The third explanation rests on differences in the experimental procedures that the groups underwent prior to providing presence ratings. It is this explanation which is more problematic for subjective presence evaluations, as the problem resides not in the rating procedure but rather in the observers not understanding the presence construct. It is possible that observers build a model of what the experimenter might mean by *presence* based on task demands and available experimental cues. For Experiment 1, for example, presence might be heavily weighted on viewing condition because of the recruitment procedure, prior stereopsis testing, and the most salient feature of the display. For Experiment 2b, presence might be heavily weighted by interest, or less heavily weighted by stereopsis (since observers had already provided interest ratings for the stimuli).

This suggestion is a concern for research utilizing direct subjective ratings of presence as it implies that the ratings may be influenced by the experimental context in which they were obtained. In an extreme case, observers may identify presence with a salient feature to which their attention is drawn in the training phase of the experiment. In a less extreme case, the interpretation of presence may be biased towards this feature. In either

case, procedures and instructions should be adapted to minimize this bias.

Experiment 3 directly tests the hypothesis that presence ratings for stimuli can be affected by previous experience.

4 Experiment 3—Effects of Practice on Subjective Presence Ratings

Three groups of observers were asked to provide ratings for the training stimulus and experimental stimulus sequences. One group rated the training stimulus for presence, one for interest, and one for three-dimensionality. After this, observers from all three groups rated the experimental sequences for presence.

If ratings of presence are assimilated toward previously rated attributes (the explanation we have advanced for the difference between the results of Experiment 1 and 2b), then we predict that the effect of viewing condition will not be as strong for the groups that rate the practice sequences for presence and interest as it will be for the group that rates the practice sequence for three-dimensionality. For the group that rates the practice sequence for interest, this prediction is made because no significant effect of viewing condition on interest was found in Experiment 2a, a result that appears to have affected their subsequent ratings of presence. For the group that rates the practice sequence for presence, the prediction is made because of the methodological changes instituted after Experiment 1, namely that the stereoacuity test is given after the experiment and that advertisements for participation are billed “new media research” rather than “3DTV research.” With the elimination of these potential sources of bias, we predict that the three-dimensionality of the stimulus will have less impact on the presence ratings produced by this group. However, we predict that the effect of viewing condition produced by the group that rates the training sequences for three-dimensionality will be larger than that produced by either of the other two groups, because observers in this group are instructed to attend to the three-dimensionality of the display in their training. We also predict that there will be a significant effect of motion for this group,

because of the capacity of motion parallax to specify a 3-D spatial layout.

4.1 Method

4.1.1 Observers Three groups of 24 observers (36 male, 36 female, average age of 24 years) volunteered to participate in the experiment. All had normal or corrected-to-normal vision. All observers had a stereoacuity of 30 sec-arc or better.

4.1.2 Apparatus and Procedure These were again identical to those in the previous experiments.

Observers in all three groups provided continuous subjective presence ratings using the continuous assessment methodology already described. All observers saw the same 3 min. training film and experimental sequences as used for Experiment 1 and 2. The only difference in the procedure undertaken by each of the three groups was the stimulus attribute on which they rated the training stimulus. One group rated the training stimulus for presence, one for interest, and one for three-dimensionality. After the data had been collected, each observer was tested for stereoscopic vision.

4.2 Results and Discussion

Figure 5 shows the group mean raw presence ratings for each of the three training groups, (presence, interest, and three-dimensionality, respectively) split by viewing condition and pooled across all the motion sequences. Across all the training groups, stereoscopic presentation resulted in higher presence ratings than did monoscopic presentation.

A three-factor mixed-measures ANOVA (training group \times viewing \times motion) was run with input data being each observer's average raw score across each motion sequence. There was a significant main effect of viewing condition ($F(1, 69) = 11.521, p < 0.001$), whereby observers produced higher presence ratings for the sequences when they were presented stereoscopically than when they were presented monoscopically. The presence-enhancing capacity of stereoscopic viewing was robust across the independent training groups, although

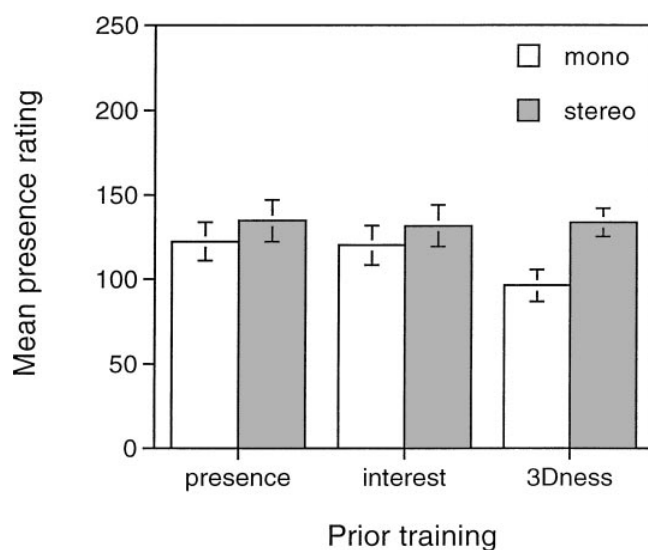


Figure 5. Data from Experiment 3. Interaction of training group and viewing condition pooled across motion-parallax levels on group mean subjective ratings of presence. Bars indicate standard errors.

a more detailed analysis reveals that this effect was magnified for the group trained on three-dimensionality, as can be seen in Figure 5. This result was as predicted and an a priori contrast showed that the difference between monoscopic and stereoscopic presentation for the group trained on three-dimensionality was significantly higher than for the groups trained on presence and interest ($t = -1.997$, $df = 69$, $p < 0.05$, two-tailed). Further examination of the graph reveals that the larger difference was caused by observers in this group rating monoscopically presented stimuli to be of lower presence than did observers from the other two groups.

In addition, there were no main effects of motion ($F(2, 138) = 0.065$, n.s.) or of training group ($F(2, 69) = 0.544$, n.s.), but there was a significant interaction between these two factors ($F(4, 138) = 2.629$, $p < 0.05$).

Figure 6 illustrates the interaction, showing the group mean raw presence ratings for each motion sequence, observer, scene, and minimal, by training group and pooled over viewing condition. In contrast to the other two groups, the group trained on three-dimensionality's results provided a clear relationship between motion parallax and presence, in that higher presence ratings

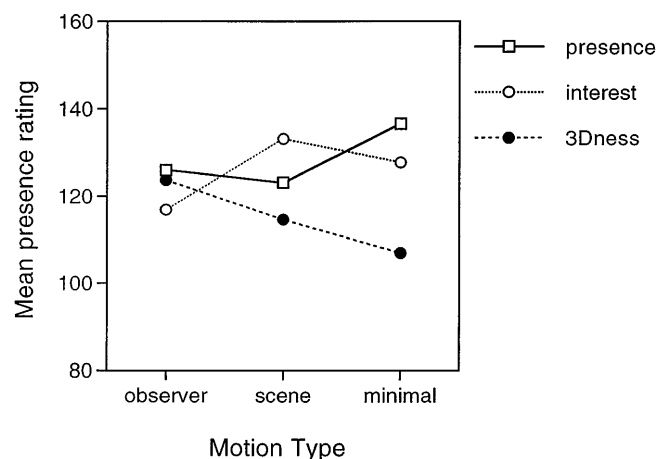


Figure 6. Data from Experiment 3. Interaction of training group and motion parallax pooled across viewing-condition levels on group mean subjective ratings of presence.

were provided for sequences containing more motion parallax, and hence more cues to 3-D depth.

The most significant result from Experiment 3 is the finding that prior training given to observers significantly affects their presence ratings. The fact that observers in the group trained on three-dimensionality rated monoscopically presented stimuli as evocative of less presence than observers in the other two groups clearly supports the hypothesis that naive observers can be influenced in their responses by their experience immediately prior to providing presence ratings. In giving observers 3 min. of practice rating three-dimensionality, the contribution of stereoscopic presentation (or three-dimensionality) to a sense of presence was magnified for them. This is evidenced by their providing lower presence ratings than the other two groups for monoscopically presented (non-3D) stimuli. This result suggests an explanation for the differences previously reported between the results of Experiment 1 and Experiment 2b.

5 General Discussion

A number of experiments have been published reporting consistent results using post-test subjective rating scales for presence (e.g., Slater and Usoh, 1994; Slater et al., 1994; Hendrix & Barfield, 1996a, 1996b).

Our own previous research has also indicated consistent results in two laboratories in different countries when the procedure was kept constant (IJsselstein et al., 1998; Freeman et al., 1997a). However, the results we have presented here demonstrate the potentially significant effects of procedure on subjective ratings of presence. Other constructs may be liable to this same methodological problem, and one such example is given by Barfield and Danas (1996). In a paper discussing the use of olfactory cues in virtual environments, Barfield and Danas state that “some of the most serious obstacles to the identification of olfaction dimensions are the fact that there exists no accepted reliable verbal classification scheme and that there is a lack of a universally endorsed system for odour classification.” While the problem is not identical—for olfaction the problem is one of classification whereas for presence the problem is one of magnitude description—the underlying problem is the same: it is that in our lexicon we do not have adequate means of describing either of the subjective sensations. It is specifically because of this fact that subjective presence evaluation is potentially unstable at present. Because observers picked at random from a population are unlikely to have experienced and discussed virtual reality systems with presence-evoking capacity, it is unreasonable to expect them to rate environments or systems without a clear explanation of the nature of presence or experience at the task. It is thus understandable that the context within which naive observers provide their ratings (pre- or mid-test variables they are asked to rate, or the most salient feature of a display) might be interpreted as indicators of components of presence, and hence influence the ratings obtained.

Our results show that a simple rating procedure for presence can be biased by previous ratings, suggesting that different attributes of the display may load more heavily on presence ratings. To avoid this pitfall we need either a bias-free method of measuring presence, or a more extensive training procedure to make clear the subjective dimension on which observers are making their assessment.

Our results also have implications for the designs of future experiments utilizing subjective assessment of presence. If studies require the measurement of different

attributes in addition to presence, then these ratings should be obtained from independent-groups designs to minimize the potential of bias affecting the measure of presence.

6 Conclusion

Our results suggest that both 3-D space perception and content-related attributes such as interest contribute to presence, but that the weighting they receive is variable by different training procedures. This suggests that direct subjective assessment of presence in naive observers is potentially unstable and subject to prior experience and task expectations. This does not mean that subjective assessment of presence is not useful, but that its limitations should be recognized and that care should be taken to avoid such bias when designing assessment procedures.

We are following two new paths of research in an attempt to improve the assessment of presence. First, we are developing objective methodologies in a search for corroboration of subjective ratings from an objective methodology. Our second line of investigation has used “focus group” methodology. Focus groups are moderated discussion groups, with six to eight participants, following a predetermined discussion guideline. The moderator ensures that progress is made through the discussion guideline. We plan to present these new displays to volunteers and ask them to discuss how the displays make them feel. From this, we aim to derive rating scales using terminology that is easily understood by nonexpert observers. The results we have presented here are thus both supportive of published theoretical discussions of the determinants of presence and challenging to studies that have used direct methods for the subjective assessment of presence.

Acknowledgements

This work was funded by the EC ACTS TAPESTRIES project and the UK Independent Television Commission. Primary grant-holder at the University of Essex is Professor Don Pear-

son of the Department of Electronic Systems Engineering (ESE). Technical assistance from Mr. Kwee Teck Tan in the Department of ESE is hereby acknowledged.

The authors also acknowledge with appreciation their partners in the EC ACTS TAPESTRIES project, particularly Dr. Huib de Ridder and Dr. Roelof Hamberg of the IPO, Center for Research on User-System Interaction (Eindhoven, NL). In addition, the authors are grateful for useful discussions shared with Professor Ray Meddis and Professor Arnold Wilkins (Department of Psychology, University of Essex) and Professor Jules Davidoff (Department of Psychology, Goldsmiths College, University of London), and for the useful and thought-provoking comments received from the reviewers of an earlier draft of this paper.

References

- Aldridge, R., Davidoff, J., Ghanbari, M., Hands, D., & Pearson, D. (1995). Measurement of scene-dependent quality variations in digitally-coded television pictures. *IEE Proceedings on Vision, Image and Signal Processing*, 142, 149–154.
- Barfield, W., Zeltzer, D., Sheridan, T. B., & Slater, M. (1995). Presence and Performance Within Virtual Environments. In W. Barfield & T. A. Furness III (Eds.), *Virtual Environments and Advanced Interface Design*. Oxford: Oxford University Press.
- Barfield, W., & Danas, E. (1996). Comments on the use of olfactory displays for virtual environments. *Presence: Teleoperators and Virtual Environments*, 5, 109–121.
- de Ridder, H., & Hamberg, R. (1997). Continuous assessment of image quality. *SMPTE Journal*, 106, 123–128.
- 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, 247–259.
- Freeman, J., Avons, S. E., & Davidoff, J. (1997a, April). *Effects of stereo and motion manipulations on measured presence in stereoscopic displays*. Paper presented at Applied Vision Association AGM, University of Abertay.
- Freeman, J., Avons, S. E., Davidoff, J., & Pearson, D. E. (1997b). Effects of stereo and motion manipulations on measured presence in stereoscopic displays. *Perception*, 26 (suppl.), 42.
- Hamberg, R., & de Ridder, H. (1995). Continuous assessment of perceptual image quality. *Journal of the Optical Society of America*, 12, 2573–2577.
- Heeter, C. (1992). Being there: the subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1, 262–271.
- Hendrix, C., & Barfield, W. (1996a). Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, 5, 274–289.
- Hendrix, C., & Barfield, W. (1996b). The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments*, 5, 290–301.
- IJsselstein, W., de Ridder, H., Hamberg, R., Bouwhuis, D., & Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. *Displays*, 18, 207–214.
- IJsselstein, W., Freeman, J., Avons, S. E., Davidoff, J., de Ridder, H., & Hamberg, R. (1997). Continuous assessment of presence in stereoscopic displays. *Perception* 26 (Suppl.) 42–43.
- ITU-R Recommendation BT.500-7 (REVISED) (1995). Methodology for the subjective assessment of the quality of television pictures.
- Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, 22, 261–270.
- Sheridan, T. B. (1992a). Defining our terms. *Presence: Teleoperators and Virtual Environments*, 1, 272–274.
- Sheridan, T. B. (1992b). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1, 120–125.
- Sheridan, T. B. (1996). Further musings on the psychophysics of presence. *Presence: Teleoperators and Virtual Environments*, 5, 241–246.
- Slater, M., & Usoh, M. (1994). Representations systems, perceptual position, and presence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 2, 221–233.
- Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 3, 130–144.
- Welch, R. B., Blackmon, 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*, 5, 263–273.
- Zeltzer, D. (1992). Autonomy, interaction and presence. *Presence: Teleoperators and Virtual Environments*, 1, 127–132.