Katerina Mania
k.mania@sussex.ac.uk
School of Engineering and Information Technology
University of Sussex
Falmer, Brighton, BN1 9QT,
United Kingdom

Tom Troscianko
University of Bristol,
United Kingdom

Rycharde Hawkes
Hewlett Packard Laboratories,
United Kingdom

Alan Chalmers
University of Bristol,
United Kingdom

Fidelity Metrics for Virtual Environment Simulations Based on Spatial Memory Awareness States

Abstract

This paper describes a methodology based on human judgments of memory awareness states for assessing the simulation fidelity of a virtual environment (VE) in relation to its real scene counterpart. To demonstrate the distinction between task performance-based approaches and additional human evaluation of cognitive awareness states, a photorealistic VE was created. Resulting scenes displayed on a head-mounted display (HMD) with or without head tracking and desktop monitor were then compared to the real-world task situation they represented, investigating spatial memory after exposure. Participants described how they completed their spatial recollections by selecting one of four choices of awareness states after retrieval in an initial test and a retention test a week after exposure to the environment. These reflected the level of visual mental imagery involved during retrieval, the familiarity of the recollection and also included guesses, even if informed. Experimental results revealed variations in the distribution of participants' awareness states across conditions while, in certain cases, task performance failed to reveal any. Experimental conditions that incorporated head tracking were not associated with visually induced recollections. Generally, simulation of task performance does not necessarily lead to simulation of the awareness states involved when completing a memory task. The general premise of this research focuses on how tasks are achieved, rather than only on what is achieved. The extent to which judgments of human memory recall, memory awareness states, and presence in the physical and VE are similar provides a fidelity metric of the simulation in question.

1. Introduction

The mapping from the real-world environment to the computer graphics environment is mediated by environmental or visual fidelity (Waller, Hunt, & Knapp, 1998). The term *visual fidelity* refers to the degree to which visual features in the virtual environment (VE) conform to visual features in the real environment. *Interface* or *interaction fidelity* refers to the degree to which the simulator technology (visual and motor) is perceived by a trainee to duplicate the operational equipment and the actual task situation. It is argued that training, for instance, in a VE with maximum fidelity would result in transfer equivalent to real-world training because the two environments would be indistinguishable (Waller et al., 1998). Robust metrics are essential to assess the fidelity of VE implementations comprising computer graphics imagery, display

Presence, *Vol. 12, No. 3, June 2003,* 296–310

technologies, and 3D interaction metaphors across a range of application fields. Apart from optimization of technological characteristics such as resolution, field of view (FOV), and latency, one common belief is that efficient task performance measures should serve as fidelity metrics for any application that mainly targets transfer of training in the real world (Bailey & Witmer, 1994; Waller et al., 1998; Lathrop & Kaiser, 2002). A commonly employed strategy, therefore, for assessing the simulation fidelity of a VE is to compare task performance in a VE to task performance in the real-world scene represented in the VE. Another common approach is to employ a cross-application construct, such as the sense of "presence" to assess the effectiveness of a VE or aspects of a VE according to its success in enhancing presence. There is a widespread belief that presence should somehow improve task performance, although this has yet to be verified or indeed reasons offered as to why this should be the case (Stanney et al., 1998).

This paper argues that, because of the wide range of VE applications and differences in participants across their background, ability, and method of processing information, an understanding of how tasks are undertaken within a VE complementing what is achieved, is significant. This rationale is applied here to spatial memory recall. The utility of VEs, regardless of the applications they are proposed for is predicated upon the accuracy of the spatial representation formed in the VE. The framework to be presented has been drawn from traditional memory research adjusted to form an experimental procedure to compare real scenes and their computer graphics simulated counterparts. Here, participants could describe how they achieved their spatial recollections after exposure to an environment by selecting one of four awareness states ("remember," "know," "familiar," or "guess") (Tulving, 1985, 1993; Conway, Gardiner, Perfect, Anderson, & Cohen, 1997; Gardiner, 2000). These judgments reflect the level of visual mental imagery involved at retrieval and the familiarity of the recollection including guesses, even if informed. To demonstrate the varied distribution of cognitive activity even when task performance remains the same, a photorealistic VE was created displayed on a head-mounted display (HMD)—incorporating either mono or stereo rendering with or without head tracking—and desktop display. Resulting scenes were then compared to the real-world task situation they represented, employing memory recall of elements of the space as well as report of awareness states on an initial test and a retention test a week after the initial exposure. Central to this work is identifying whether experimental conditions such as the real-world one and those incorporating head tracking (thus including proprioceptive information) are associated with stronger visually induced recollections ("remember" awareness state) compared to conditions associated with a typical mouse interaction interface. This work also aims to explore whether a cognitive shift between initial test and retest is going to signify a performance shift. This study extends a preliminary study by Mania and Chalmers (2001).

2 Background

2.1 Spatial Training in Synthetic Worlds

The first effort to compare real and simulated computer graphics static scenes side by side was attempted by Meyer, Rushmeier, Cohen, and Greenberg (1986). Radiometric values predicted using a radiosity rendering of a basic scene were compared to physical measurements of radiant flux densities in the real scene, both of which were viewed through the back of a view camera. In a more recent approach, McNamara, Chalmers, Troscianko, and Gilchrist (2000) described a method for measuring the perceptual equivalence between a real scene and static computer simulations of the same scene based on human judgments of lightness. Results showed that rendering solutions such as tone mapping were of the same perceptual quality as a photograph of the real scene.

For real-time VE applications, a central research issue for training could be how participants mentally represent an interactive computer graphics world and how their recognition and memory of such worlds correspond to actual conditions. Waller et al. (1998), Bailey and Witmer (1994), and Bliss, Tidwell, and Guest (1997) examined the variables that communicate transfer of spatial knowledge and discuss the form and development of spatial representation in VE training in relation to either real-world training or training with maps, photographs, and blueprints. The suitability of VE systems as effective training mediums is examined and concluded to be as effective as map or blueprint training (Waller et al., 1998; Bliss et al., 1997) with configurational knowledge acquisition similar to training with photographs and real-world training (Bailey & Witmer, 1994). Dinh, Walker and Hodges (1999) investigated the effects of tactile, olfactory, audio, and visual sensory cues on participants' memory recall of a building. Two levels of visual detail were investigated by reducing texture resolution with or without ambient auditory, olfactory, and tactile stimulation. No significant main effect was revealed on spatial layout recall. Accurate recall of objects' locations was significantly higher when tactile and olfactory cues were incorporated in their environment. Arthur, Hancock, and Chrysler (1997) examined participants' ability to reproduce a complex spatial layout of objects having experienced them previously under different viewing conditions (a free binocular virtual condition, a free binocular real-world condition, and in a single-viewpoint monocular view of the real world). Mapping results showed a significant effect of viewing condition where, interestingly, the static monocular condition was superior to both the active virtual and real binocular conditions.

Experimental postexposure methodologies for spatial recall investigation range from questionnaires (Dinh et al., 1999) to drawing sketches of a space after exposure (Billinghurst & Weghorst, 1995) or applying the spatial knowledge acquired so as to navigate effectively the real-world space represented (Waller, et al., 1998; Bailey & Witmer, 1994; Bliss et al., 1997). Performance accuracy is the dominant means of assessing a VE simulation. In this paper, performance accuracy is complemented by self-report of awareness states during retrieval, utilizing a memory task and a retention memory test a week after exposure across real-world and photorealistic VE viewing conditions.

2.2 Memory Awareness States Methodology

Memory, in the sense of "information" for subsequent analysis, plays an important role in perceptual systems such as the visual, auditory, haptic, and kinesthetic. Memory is not a unitary system (Baddeley, 1997): in the process of acquiring a new knowledge domain, visual or nonvisual, information retained is open to a number of different states. Some elements of a learning experience or of a visual space may be "remembered" linked to a specific recollection event and mental image or could just popout, thus, could be just "known." According to Tulving (1985), recollective experiences are the hallmark of the episodic memory system. Knowing refers to those in which there is no awareness of reliving any particular events or experiences, a mental thesaurus (semantic memory). Tulving introduced a distinction between *remember* and *know* responses and provided the first demonstration that these responses can be made in a memory test, item by item out of a set of memory recall questions, to report awareness states as well. He reported illustrative experiments in which participants were instructed to report their states of awareness at the time they recalled or recognized words they had previously encountered in a study list. If they remembered what they experienced at the time they encountered the word, they made a "remember" response. If they were aware they had encountered the word in the study list but did not remember anything they experienced at that time, they expressed a "know" response. The results indicated that participants could quite easily distinguish between experiences of remembering and knowing.

There is some preliminary evidence that the distinction between remembering and knowing reflects a difference in brain activity at the time of encoding (Smith, 1992). It is assumed that recognition memory can be based largely on knowing, with little or no remembering. All that is necessary for encoding into the semantic system is some initial awareness of events. In contrast, encoding into episodic memory must depend on greater conscious elaboration of the events. Gregg and Gardiner (1994) showed that estimates of the strength of the memory trace are greater when derived from remember plus know responses than when derived from only remember responses. Knowing, thus, reflects an additional source of memory, not merely a difference in response criteria. Although, remember and know awareness states have been controversially linked to episodic and semantic memory types with know responses more theoretically problematic, recent research emphasized that “they can be used without commitment to any theory, but simply to provide information on how various phenomena, including memory disorders, are characterised experientially” (p. 933) (Gardiner, 2000). In a relevant study, overall recognition performance in two groups of participants was very similar; however, the reported states of awareness differed markedly. One cannot make assumptions on what participants experience mentally from only their performance; therefore, there is no alternative to the use of subjective reports. Thus, additional information of awareness states provides an invaluable input into how participants complete recollections. Subsequent research to Tulving (1985), summarized in Gardiner (2000), demonstrated that some variables affect one or the other of the two states of awareness, that some variables have opposing effects on them, and that some variables have parallel effects on them. This finding indicates that the two states of awareness are functionally independent.

Conway et al. (1997) argued that *familiarity* can be defined as the feeling that something has been encountered or experienced recently, although nothing about this recent occurrence can be remembered. Know responses, on the other hand, represent highly familiar memory items that may come to mind without recollecting any particular encounter or any feeling of a recent encounter and cannot be placed. Conway et al. showed that these finer-grained judgments could be dissociated from each other, just as different source memory judgments can. A confidence scale cannot communicate awareness states. It is also suggested that, when a new knowledge domain is to be acquired, memory is represented initially in an episodic way. As time goes by, the underlying representations may change such that they do not represent recollective experiences and are simply known, leading to a semantic representation and schematized conceptual knowledge. There is little evidence that feelings of familiarity reflect the semantic memory system that supports highly familiar long term knowledge. Gardiner (2000) concludes: “psychology of memory should take on board subjective reports of conscious states and not just rely on more conventional measures of performance. This evidence has established that the essential subjectivity of remembering and knowing does not make reports of these states of awareness intractable to science” (p. 940).

3 Experimental Methodology

3.1 Experimental Design

Five groups of 21 participants were recruited from the University of Bristol undergraduate and M.Sc. student population, and they received course credits for their participation. Eighty percent of the participants from each group were men, and all used computers a great deal in their daily activities. A between-subject design was utilized, balancing groups for age and gender. Participants in all conditions were informed that they could withdraw from participation at any time during the experiments and they were naive as to the purpose of the experiment. Participants had either normal or corrected-to-normal vision (self-report). According to the group they were assigned to, participants completed the same memory task in one of the following conditions.

1) In reality, wearing custom-made goggles to restrict their FoV, allowing for monocular vision; referred to as the *real-world condition*.
2) Using a photorealistic computer graphics simulation on a monocular head-tracked HMD; referred to as the *HMD mono head-tracked condition*.
3) Using the same application on a stereo head-tracked HMD; referred to as the *HMD stereo head-tracked condition*.
4) Using the same application on a monocular HMD with a mouse interface; referred to as the *HMD mono mouse condition*.
5) Using the same application displayed on a typical

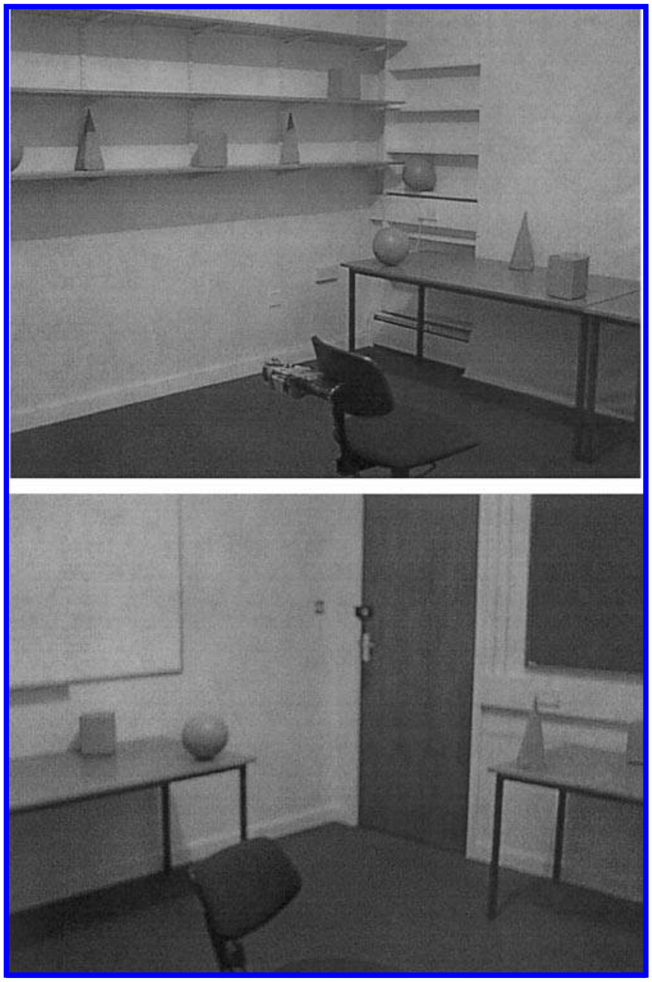

Figure 1. *The real-world room (real-world condition).*

desktop monitor with a mouse interface, wearing the same restrictive goggles as in the real-world condition; referred to as the *desktop condition*.

A week after their experience, all participants were retested on the same memory task.

3.1.1 The Real Environment. The real environment consisted of a 4 × 4 m room (Figure 1). Each wall of this room had a different landmark: one wall consisted of a door and shelves, one wall a door and a greenboard, the third wall a whiteboard, and the fourth smaller shelves on both its ends. The existing window in the room was firmly covered with black lining to keep natural light out. The light fixtures in the room were replaced with a standard incandescent bulb (assumed diffuse, light emission in all directions). Several tables were placed close to the walls, and 21 primitive objects of approximately the same size (seven boxes, seven spheres, and seven pyramids) were scattered around the room, on the tables and shelves. All the objects were painted one shade of blue using the same diffuse paint. A swivel chair was placed in the middle of the room.

3.1.2 The Computer Graphics Simulation. There was tight control over the visual appearance of the experimental space across real-world and simulated conditions. The geometry in the real room was measured using a regular tape measure with accuracy of the order of 1 cm. A photometry instrument (Minolta Spot Chromameter CS-100) was used to measure the chromaticity CIE(x,y) and luminance (Y) values of the light and materials in the real room. The Minolta chromameter is a compact, tristimulus colorimeter for noncontact measurements of light sources or reflective surfaces. Luminance relates to the quality of a color that most resembles the human's notion of brightness. Bright colors are generally of a high luminance and dark colors are generally of a low luminance. The illuminant (light source) was measured by placing a white sheet of paper in a specific position. Most of the materials (walls, objects, shelves, floor, plugframes) were measured at the same position. To ensure accuracy, five measurements were recorded for each material, the highest and lowest luminance magnitudes were discarded, and an average was calculated of the remaining three triplets. However, as this was a room in daily use, some variations existed in all of the surfaces due to texture, age, and dirt.

The CIE (1931) color space is based on color matching functions derived by human experimentation, and it incorporates the trichromacy of the Human Visual System (HVS). The usefulness of the CIE(x,y) representation is that it allows color specification in one language; however, equal geometric steps of CIE(x,y) space do not correspond to equal perceptual steps. Before specifying display colors, it is necessary to compute the tristimulus matrix of the display in question. To compute the RGB tristimulus matrix, the chromaticity coordinates of the three display phosphors in CIE(x,y) space are required. In addition, the chromaticity coordinates of the white that the three phosphors of the display produce when turned on at their maximum are also required (Travis, 1991). Generally, the RGB system is a means for describing colors on a display monitor. It does not take into account the energy that is produced in the physical world in terms of the distribution over wavelength and also how HVS responds to this distribution.

For the final measurements, the illuminant had to be taken into account. Measuring a diffuse surface under a given light source results in Yxy values, which include the contribution of the light source itself. Incandescent bulbs are quite orange, and fluorescent light is quite green; however, the HVS perceives light in relative values and not as absolute measurements such as the ones out of the chromameter. For example, if 1000 is the luminance in the real world, 100 the luminance of a real-world material, but 100 the luminance in the computer graphics simulation, then the luminance for the simulated material needs to be 10 for the same ratio to be preserved. The color constancy attribute of the HVS, generally, is responsible for humans perceiving a white sheet of paper as white under a wide range of illumination. If a participant is immersed into a synthetic space on a display, theoretically, this should be true as well, but the small size of the displays prevents color constancy from occurring. In relevant calculations for simulating real-world illumination in a synthetic world, therefore, color constancy needs to be enforced in the rendering process because the HVS does not function as in the real world due to the nature of the displays. The color of the illuminant in RGB values was set as (1,1,1) for the radiosity rendering white.

To render the scene, the materials' diffuse color needs to be specified, not the color observed under a particular light source. The final color for each measured material in the scene is estimated by dividing its RGB value by the RGB value of the observed white in the scene, which is the color of the light source in the scene. Using the relevant geometry and surfaces and illuminant measurements converted to RGB triplets as input, the rendered model was created using a radiosity rendering system (Figure 2). The final radiosity solution consisted of a finely meshed model that could be interactively manipulated. This was the basis for the application displayed on the desktop monitor and on the HMD. The desktop monitor and the HMD were gamma corrected using the Minolta Spot Chromameter CS-100 to acquire relevant luminance readings. When accurate colour specification is required as is often the case in scientific applications, the non-linear relationship between display luminance and voltage is a significant source of error and needs to be corrected to linearity.

Figure 2. *The radiosity rendering.*

3.2 Materials

The five groups of participants were asked to complete the same set of questionnaires. This set included the SSQ questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) before and after the task, the memory task and memory awareness states questionnaire, and the presence questionnaire (Slater, Steed, McCarthy, & Maringelli, 1998). All participants across the five conditions completed the same memory task a week after the initial experiment reporting on memory recall, confidence, and awareness states.

3.2.1 Memory Recall Task. The memory recall questionnaire was designed to test the participants'

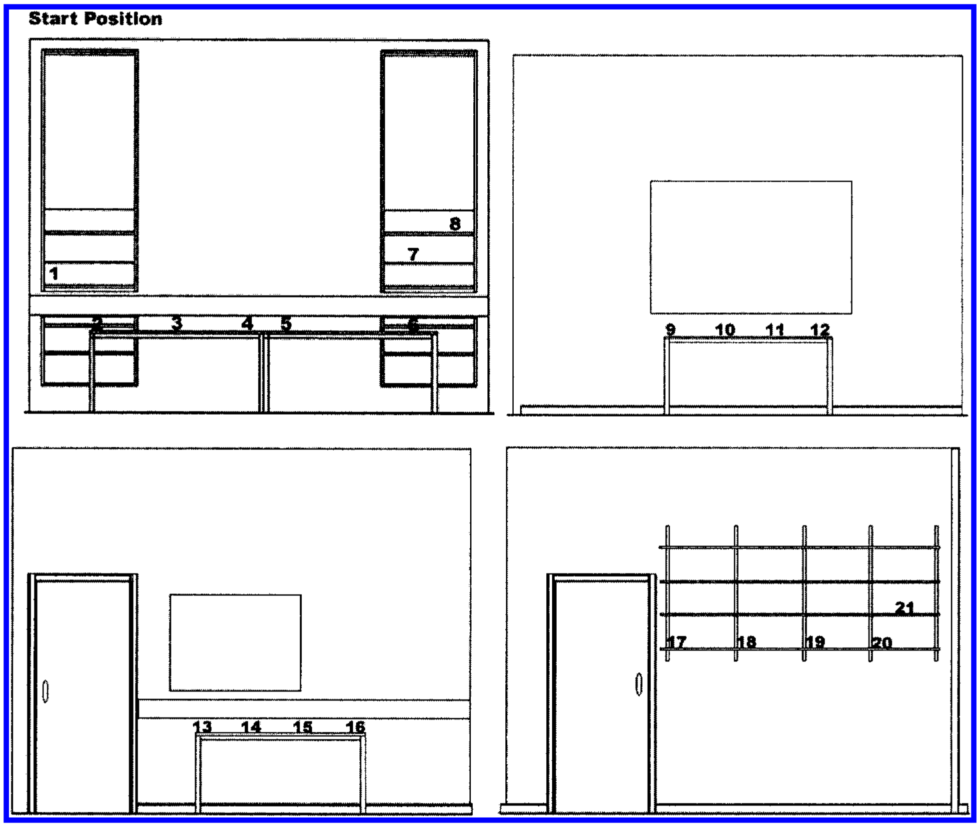

Figure 3. *Diagrams utilized for memory recall testing, for each wall of the room. (Included here in high and low resolution)*

memory recall of the positions and geometric shape of the 21 objects in the room. A diagram for each wall in the room included numbered positions of objects in various locations. The diagrams were administered together with the task questionnaire, which consisted of 21 multiple-choice questions representing the 21 objects in the scene (Figure 3). Every question included three possible answers (box, sphere, or pyramid) and a confidence scale with five possible states: no confidence, low confidence, moderate confidence, confident, certain. Every question also included an awareness states report for every recollection, based on the memory awareness methodology offering four choices: remember, know, familiar, or guess. The participants were required to report on the shape of the object in each numbered position on the diagram, starting with the positions they were more confident that they remembered. The design, thus, of the task questionnaire did not force participants to start from a specified position in the room offering the capability to report, initially, their most confident recollections. A pilot study was conducted to determine the number of objects and, therefore, the number of questions of recall in relation to the exposure time so as to avoid possible floor or ceiling effects (the task being too easy or too hard). Prior to filling out the core of the task questionnaire, participants were given instructions designed to explain what the memory awareness states depicted, as follows.

- REMEMBER means that you can visualize clearly the object in the room in your head, in that particular location. You virtually “see” again elements of the room in your mind.
- KNOW means that you just “know” the correct

answer and the alternative you have selected just "stood out" from the choices available. In this case you can't visualize the specific image or information in your mind.

- FAMILIAR means that you did not remember a specific instance, nor do you know the answer. It may seem or feel more familiar than any of the other alternatives.
- GUESS means that you may not have remembered, known, or felt that the choice you selected had been familiar. You may have made a guess, possibly an informed guess, e.g., you have selected the one that looks least unlikely.

3.2.2 Other Measures. The presence questionnaire developed by Slater et al. (1998) was designed to measure the level of presence on a Likert seven-point scale and was administered after the initial memory recall task across conditions. The widely used Simulator Sickness Questionnaire (SSQ) was administered before and following participants' exposure across conditions (Kennedy et al., 1993).

3.3 Procedures

3.3.1 The Real-World Condition. The SSQ was administered before exposure. Following this procedure, participants were asked to wear any glasses or contact lenses they normally use when they have to focus at 2 m distance (self-report). Subsequently, their dominant eye was identified by a widely used sighting test. A predetermined viewing position was set by manipulating the height of the swivel chair according to the individual. Appropriate goggles were worn that restricted participants' FOV to 30 deg. to match the desktop and HMD's FOV allowing for monocular vision through the dominant eye only (Figure 4). The FOV was restricted in the real-world condition to match the FOV of the displays. Although this action resulted in a "window" to the real world through the goggles, it was considered necessary to keep the FOV constant across conditions. Participants were instructed that they would be guided to a room where they would spend 3 min. observing by rotating the swivel chair they would sit on placed in the middle of the room; however, they were not aware of the postexposure task. Navigational patterns and idle time were monitored and recorded during exposure through a digital compass attached on the swivel chair (Mania & Randell, 2002). After the set exposure time of 3 min., participants were guided to the test room where the questionnaire pack was administered together with the appropriate instructions.

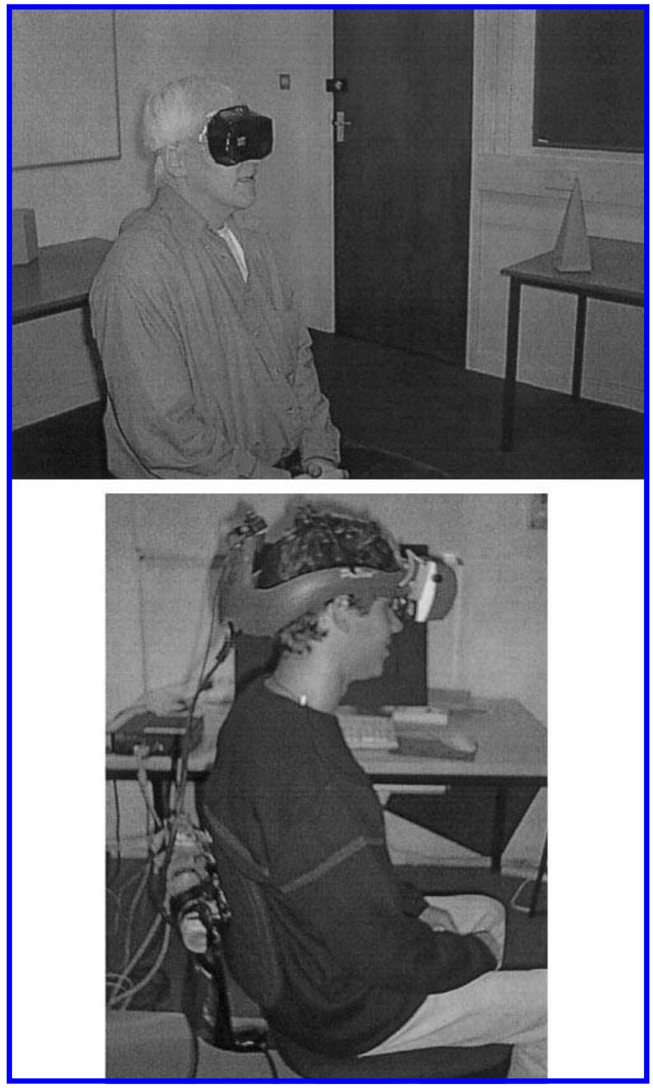

Figure 4. *The real-world and HMD mono/stereo condition (head-tracked).*

3.3.2 The Display Conditions. The computer graphics application was displayed on a Kaiser Pro-View

30, gamma-corrected HMD (Figure 4). The viewpoint was set in the middle of the room and navigation was restricted to a 360° circle around that viewpoint and 180° vertically to simulate participants' movement on the swivel chair in the real room (three degrees of freedom). The geometric FOV was calculated to be the same as the visual angle, through the goggles, in the real room. For the HMD monocular conditions (head-tracked and non-head-tracked), the dominant eye was identified and the appropriate screen of the HMD was covered, allowing for vision through only the dominant eye. For the HMD stereo head-tracked condition, each participant's interpupilary distance (IPD) was measured, and the stereo application's parallax was set accordingly for the individual. For the desktop condition utilizing a gamma-corrected, typical 21 in. desktop monitor, each participant's dominant eye was identified, and the appropriate goggles were subsequently worn as in the real-world condition. The frame of the monitor was covered with black cardboard to achieve a foreground occlusion effect, resulting in a stronger sense of depth. Horizontal rotation was monitored across all conditions (Mania & Randell, 2002). There was no other source of light besides the HMD or desktop display during exposure. The frame rate was retained at fourteen frames per second across all conditions. Although this is not a particularly high frame rate, it was considered adequate. The display resolution was 640 × 480 (HMD maximum resolution) across technological conditions and the FOV was constant (30°) across all conditions, including the real-world condition with restrictive goggles fitted. The computer graphics rendering was computed taking into account real-world photometric measurements resulting in a photorealistic rendering as described in the previous section. Texture mapping was applied only on the doors and tables in the room.

4 Results and Discussion

4.1 Memory Awareness States' Statistical Analysis

Awareness state data were represented as prior and posterior probabilities. Koriat and Goldsmith (1994) have drawn an important distinction between the amount or quantity remembered compared to the accuracy or quality of what is remembered. In the quantity analysis memory awareness states, data are represented as a priori or prior probabilities. Although this notation does not follow the Bayesian probability theory principles for prior probabilities, it is going to be adopted as such in this paper following the characterizations of Koriat and Goldsmith (1994) as well as Conway et al. (1997). Prior probabilities are obtained by calculating the proportions of correct answers falling in each of the four memory awareness categories for each participant. In the accuracy analysis, correct recall scores are represented as posteriori or posterior probabilities. To calculate posterior probabilities, the proportion of correct answers from the total of answers given in each memory awareness category is computed for each participant.

For participant n,

X_{in} is the number of correct answers for the i awareness state,
X'_{in} is the number of incorrect answers for the i awareness state,
i = {remember, know, familiar, guess} = {1,2,3,4}

then,

P_{in} is the prior probability for awareness state i related to participant n,

$$P_{in} = \frac{X_{in}}{\sum_{i=1}^{4} X_{in}}, \tag{1}$$

and

P'_{in} is the posterior probability for awareness state i related to participant n,

$$P'_{in} = \frac{X_{in}}{X_{in} + X'_{in}}. \tag{2}$$

Generally, prior probabilities pose the following question: given that the response of a participant is correct, what is the probability that the participant has chosen a particular state on that question? Posterior probabilities,

Table 1. *Means and Standard Deviations for Accurate Memory Recall Performance and Confidence Scores as a Function of Viewing Condition (N total number of participants)*

Viewing condition	Recall performance		Confidence	
N = 105	Initial test	Retest	Initial test	Retest
Real world	12.42 (4.29)	10.14 (4.60)	3.22 (0.78)	2.50 (0.77)
HMD mono head-tracked	12.25 (5.05)	10.75 (4.41)	2.93 (0.85)	2.05 (0.86)
HMD stereo head-tracked	10.89 (3.63)	9.42 (3.80)	2.66 (0.63)	1.94 (0.83)
Desktop	10.90 (3.94)	8.00 (3.53)	2.94 (0.55)	2.27 (0.64)
HMD mono mouse	10.56 (3.21)	8.34 (3.21)	2.99 (0.60)	2.15 (0.73)

on the other hand, ask: given that a response of a participant was assigned to one of the four memory awareness response categories, what is the probability that the response is correct? For the purpose of this study each memory recall question included a five-point confidence scale and a choice between remember, know, familiar, and guess awareness states. The goal of this strategy was to identify the distributions of awareness states responses across conditions focusing on visually induced recollections. This could reveal variations that wouldn't be possible by just counting right and wrong answers.

4.2 Spatial Memory Recall and Memory Awareness States Results

The participants completed the memory task including confidence and awareness responses across the five conditions. The memory recall scores for the initial task and retest, the confidence scores as well as the prior and posterior probabilities derived from the memory awareness states data set were analyzed using ANOVA, a powerful set of procedures used for testing significance when two or more conditions are used. Significance decisions involve rejecting or retaining the null hypothesis (which claims that groups are identical). The null hypothesis is rejected when the probability that a result occurring under it is less than .05 (Coolican, 1999).

The total number of objects that were correctly located and identified was counted for each participant after completing the initial test as well as the retention test a week after. The memory performance measures were subjected to a 5 (viewing condition) × 2 (testing session) mixed ANOVA with viewing condition as a between-subjects factor and testing session as a within-subjects factor, with number of correct responses as the dependent variable. Table 1 shows the mean accurate recall scores and standard deviations (in parenthesis) as a function of viewing condition and test/retest session. All effects were evaluated at a p level of .05 to determine statistical significance. There was a significant main effect for testing session ($F(1,100) = 36.51$, $p < .01$) but not for viewing condition ($F(4,100) = 1.47$, $p > .05$). The interaction between testing session and viewing condition was not significant. These results show that participants had retained significantly less spatial information over time; however, the viewing condition had no effect on the decrease of recall performance.

A confidence measure was included for each recollection. The confidence scores were subjected to a 5 (viewing condition) × 2 (testing session) mixed ANOVA with viewing condition as a between-subjects factor and testing session as a within-subjects factor, with the confidence selection (no confidence = 1, low confidence = 2, moderate confidence = 3, confident = 4, certain = 5) as the dependent variable. Table 1 shows the mean confidence scores and standard deviations (in parenthesis) as a function of viewing condition and test/retest session. There was a significant main effect for testing session ($F(1,100) = 183.59$, $p < .01$), but not for view-

Table 2. *Prior Probabilities and Standard Deviations as a Function of Viewing Condition and Test/Retest Session (N total number of participants)*

Viewing condition N = 105	Remember test	Remember retest	Know test	Know retest	Familiar test	Familiar retest	Guess test	Guess retest
Real world	0.33 (0.25)	0.16 (0.25)	0.26 (0.25)	0.15 (0.24)	0.27 (0.24)	0.33 (0.28)	0.12 (0.16)	0.34 (0.29)
HMD mono head-tracked	0.27 (0.30)	0.04 (0.09)	0.25 (0.29)	0.13 (0.24)	0.24 (0.21)	0.29 (0.20)	0.22 (0.20)	0.53 (0.29)
HMD stereo head-tracked	0.24 (0.23)	0.10 (0.19)	0.16 (0.19)	0.07 (0.16)	0.39 (0.3)	0.37 (0.26)	0.18 (0.18)	0.44 (0.29)
Desktop	0.29 (0.24)	0.04 (0.09)	0.20 (0.18)	0.13 (0.24)	0.28 (0.2)	0.29 (0.2)	0.2 (0.18)	0.53 (0.29)
HMD mono mouse	0.49 (0.22)	0.09 (0.15)	0.10 (0.14)	0.20 (0.27)	0.22 (0.18)	0.27 (0.22)	0.17 (0.17)	0.41 (0.34)

ing condition. Also, the interaction between testing session and viewing condition was not significant. These results show that participants had significantly less confidence over time while completing the memory task; however, the viewing condition had no effect on the decrease of confidence. Table 2 shows the mean prior probabilities and standard deviations (in parenthesis) as a function of viewing condition and test/retest session. Prior probabilities indicate the proportion of correct answers under each memory awareness state. The prior probabilities were subjected to a 5 (viewing condition) × 4 (awareness state) × 2 (testing session) mixed ANOVA with viewing condition as a between-subjects factor and both awareness session and testing session as within-subjects factors. There was a significant main effect for awareness state ($F(3,300) = 11.17$, $p < .05$) but not for viewing condition. The interaction between awareness state and viewing condition was significant ($F(12,300) = 1.8$, $p < .05$). The interaction between awareness state and testing session was also significant ($F(3,300) = 42.4$, $p < .05$). One-way ANOVA and Tukey's post-hoc tests were applied following the significant interaction between awareness state and viewing condition separately for the initial task and retest. There was a significant main effect of condition upon the remember awareness state, $F(4,104) = 3.016$, $p < .05$, and a tendency of significance for the know awareness state, $F(4,104) = 1.913$, $p < .1$. In particular, the probability that correct responses would be linked with the remember awareness state was significantly higher for the HMD mono mouse condition compared to the HMD mono head-tracked and HMD stereo head-tracked conditions ($p < .05$). No significant effects were revealed for the retest. The three-way interaction between viewing condition, awareness state, and testing session was not significant. A thorough inspection of the prior probabilities means reveals that correct remember responses dramatically declined over time and correct guess responses substantially increased over time. A possible interpretation could be that correct remember responses were converted to correct guess responses at the retest, with correct know and familiar responses comparatively slightly changed.

Table 3 shows the mean posterior probabilities and standard deviations as a function of viewing condition and test/retest session. Posterior probabilities represent the probability that a memory recall response assigned to each of the memory awareness states is accurate. Posterior probabilities related to the familiar and guess awareness states were calculated for the retest. A small number of participants selected the remember and know awareness states resulting in posterior probabilities not being calculated reliably. The posterior probabilities were subjected to a 5 (viewing condition) × 4 (aware-

Table 3. *Posterior Probabilities and Standard Deviations as a Function of Viewing Condition and Test/Retest Session (N total number of participants)*

Viewing condition	Remember test, N = 55	Remember retest	Know test, N = 55	Know retest	Familiar test, N = 55	Familiar retest, N = 94	Guess test, N = 55	Guess retest, N = 94
Real world	0.90 (0.31)	—	0.62 (0.37)	—	0.47 (0.30)	0.44 (0.30)	0.60 (0.33)	0.41 (0.23)
HMD mono head-tracked	0.85 (0.25)	—	0.73 (0.33)	—	0.49 (0.34)	0.49 (0.27)	0.46 (0.26)	0.47 (0.10)
HMD stereo head-tracked	0.77 (0.31)	—	0.49 (0.35)	—	0.49 (0.33)	0.55 (0.19)	0.27 (0.31)	0.33 (0.21)
Desktop	0.65 (0.43)	—	0.68 (0.33)	—	0.43 (0.31)	0.46 (0.26)	0.43 (0.32)	0.32 (0.22)
HMD mono mouse	0.88 (0.19)	—	0.53 (0.37)	—	0.38 (0.32)	0.35 (0.28)	0.28 (0.26)	0.27 (0.21)

ness state) mixed ANOVA with viewing condition as a between-subjects factor and awareness session as a within-subjects factor, separately for test and retest. For the initial task, there was a significant main effect for awareness state ($F(3,150) = 19.70$, $p < .05$) but not for viewing condition ($F(4,50) = 1.20$, $p > .05$). The interaction between viewing condition and awareness state was not significant. For the retest, there was a significant main effect for awareness state ($F(1,89) = 9.47$, $p < .05$) and for viewing condition ($F(4,89) = 2.62$, $p < .05$). The interaction between viewing condition and awareness state was not significant. A subset of participants was included in this analysis due to the issues just mentioned.

Correlation analysis between the prior probabilities derived from the awareness states results and confidence scores revealed a varied pattern of significant correlations (Pearson's):

- There was a significant positive correlation between correct "remember" responses and confidence scores for the desktop ($r = 0.45$, $p < 0.05$), and HMD mono mouse ($r = 0.65$, $p < 0.001$) conditions.
- There was a significant positive correlation between correct "know" responses and confidence scores for the real ($r = 0.75$, $p < 0.001$), the HMD mono head-tracked ($r = 0.42$, $p < 0.05$), and the desktop ($r = 0.64$, $p < 0.001$) conditions.
- There was a significant negative correlation between correct "familiar" responses and confidence scores for the real ($r = -0.58$, $p < 0.01$), desktop ($r = -0.57$, $p < 0.01$), and HMD mono mouse ($r = -0.59$, $p < 0.01$) conditions.
- There was a significant negative correlation between correct "guess" response and confidence scores for the real ($r = -0.57$, $p < 0.01$), HMD mono head-tracked ($r = -0.78$, $p < 0.001$), HMD stereo head-tracked ($r = -0.61$, $p < 0.01$), and desktop ($r = -0.63$, $p < 0.01$) conditions.

Crucially, correct "remember" responses, which were significantly higher for the HMD mono mouse condition compared to the HMD head-tracked conditions, also positively correlated with confidence scores while, respectively, they did not for the HMD head-tracked conditions.

Generally, incorporating awareness states in a memory test connects memory recall with cognitive activity and forms a framework that investigates how humans mentally represent a space from a cognitive point of view rather than a task performance point of view. Such metrics could form an integral part of the significant performance efficiency measures.

4.3 Presence and Simulator Sickness Results

The presence questionnaire was administered after the initial task was completed. A binomial regression analysis was employed based on the count of high scores out of six presence questions and following the analysis discussed in the Slater et al. study. A value of 0 was assigned if the count of high scores was 0–2, and 1 if the count of high scores was 3–6. Binomial regression, generally, shows the probability of falling under one of the 0 or 1 binomial distributions. An overall effect of condition was not revealed. Similar effects of condition on presence were revealed in studies in which the validity of the questionnaire is examined (Usoh, Catena, Arman, & Slater, 2000). The questionnaire may have failed to pick up the difference across conditions or there was not any due, for instance, to the high quality of the rendering. A concrete understanding of presence, in a way that will allow formal assessments of its perceived level in experimental studies such as this one (if this is ever possible or desirable), will aid towards forming relevant conclusions. The SSQ scores (Kennedy et al., 1993) were low due to the short exposure time.

5 Discussion

This investigation focuses on the effect of different viewing conditions (direct perception of objects in a real-world setting versus perception of the computer graphics representation of this setting) on observers' attributions regarding object-location memory. Accuracy of performance per se is an imperfect reflection of the subjective experience that underlies performance in memory tasks. Accurate memory task performance can be accompanied by either a recollection of prior specific experience (remembering) or reliance on a general sense of knowing with little or no recollection of the source of this sense (knowing), including familiarity and guesses even if informed. Training in a VE system capable of perfectly simulating the real world should result in the same training effect as that in the real world. The participants who mentally visualized the room and the objects in the room during retrieval had a higher proportion of correct responses under the remember awareness state. The participants that employed mnemonic strategies based on words instead of visually retaining elements of the space reported the know awareness state, which resulted in a proportion of correct responses linked with the know awareness state. If a weaker trend of nonvisually induced recollections is employed by participants towards stronger visually induced recollections linked to the remember awareness state, it could be assumed that their mental representation of a space involved more "vivid" recollections.

There was a significant main effect of condition upon the remember awareness state. It was anticipated that the amount of correct remember responses would be higher in conditions incorporating more naturalistic interfaces such as head tracking. However, results revealed that the proportion of correct responses linked with the remember awareness state was significantly higher for the HMD mono mouse condition compared to the HMD mono head-tracked and HMD stereo head-tracked conditions (initial task). Crucially, these responses correlated positively with confidence scores. Therefore, an interface of high-simulation fidelity such as head tracking does not always correspond to visually induced memory awareness states. A similar result was revealed in a preliminary study by Mania and Chalmers (2001). If specific applications require a high amount of recollections based on visual mental imagery, a "natural" interface such as head tracking may not be appropriate. Therefore, desirable variations of awareness states for specific application purposes could be identified. It could be true, for instance, that for flight simulation applications it is crucial for trainees to achieve a high level of visually induced recollections related to instruments as opposed to feelings of familiarity of even confident recollections that are not accompanied by visual imagery. If "reality" is associated to the degree of similarity to the real-world task situation, then, in this case, the HMD mono mouse condition is not very "real." The awareness states distribution is affected by the degree of realism of the motor response. Word-based mnemonics and, generally, recollections that were not linked to visually induced recollections were identifiable by the high proportion of correct know responses. The utilization of a viewing method such as the HMD together with an "unreal" motor response such as the mouse, appeared to have prevented participants employing nonvisually induced recollections and resulted in a larger distribution of correct responses assigned to the remember awareness state. By decreasing the degree of "reality" of the motor response, participants paradoxically adopted visually induced recollections. Achieving high fidelity could incorporate the need for similar awareness states between a real-world task situation and its computer graphics simulation. Here, something less "real," therefore, less computationally expensive but more demanding because of its novelty, may restore a more naturalistic or desirable awareness state. Research could identify such issues by using methodologies that allow investigations based on awareness states responses. Additionally, a significant shift of correct remember responses in the initial task to correct guess responses in the retest was observed. This shift was observed across all conditions, and it did signify a lower number of correct recollections between initial test and retest.

The task employed in this study did not allow for free navigation around the experimental space. The FOV was restricted in the real-world setting to match the FOV of the displays for methodological reasons. Future work could include a task that would allow freedom of navigation and also a testing strategy that would incorporate transfer of training in the real world. Matching participants' performance in simulations to performance in a real-world situation does not guarantee that the cognitive activity linked with performance will be similar across the simulated conditions. Task performance scores could, therefore, be taken into account according to specific awareness states. By employing methodologies, such as the memory awareness states methodology, computer graphics and VE technology research could exploit human perceptual mechanisms towards successful applications.

Acknowledgments

This research was funded by the Hewlett Packard Laboratories External Research Program. We wish to thank the anonymous reviewers for their insightful comments that contributed to the final version of this paper.

References

Arthur, E. J., Hancock, P. A., & Chrysler, S. T. (1997). The perception of spatial layout in real and virtual worlds. *Ergonomics, 40*(1), 69–77.

Baddeley, A. (1997). *Human memory, theory and practice.* New York Psychology Press.

Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a virtual environment. *Proc. of the Human Factors & Ergonomics Society 38th Annual Meeting,* 1158–1162.

Billinghurst, M., & Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. *Proc. of Virtual Reality Annual International Symposium (VRAIS 1995),* 40–47.

Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence: Teleoperators and Virtual Environments, 6*(1), 73–86.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology, 126*(4), 393–413.

Coolican, H. (1999). *Research methods and statistics in psychology,* 3rd ed. London: Hodder & Stoughton.

Dinh, H. Q., Walker, N., & Hodges, L. F. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proc. of IEEE VR 1999,* 222–228.

Gardiner, J. M. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *Oxford Handbook on Memory.* Oxford: Oxford University Press.

Gregg, V. H., & Gardiner, J. M. (1994). Recognition memory and awareness: A large effect of study-test modalities on 'know' responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology, 6*(2), 131–147.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy oriented and quantity oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123,* 297–315.

Lathrop, W. B., & Kaiser, M. K. (2002). Perceived orientation in physical and virtual environments: Changes in perceived orientation as a function of idiothetic information available. *Presence: Teleoperators and Virtual Environments, 11*(1), 19–32.

Mania, K., & Chalmers, A. (2001). The effects of levels of immersion on presence and memory in virtual environments: A reality centred approach. *Cyberpsychology & Behavior, 4*(2), 247–264.

Mania, K., & Randell, C. (2002). Monitoring navigational strategies and idle time in real and virtual environments: An experimental study. *Proc. of the 8th International Conference on Virtual Systems and Multimedia 2003 (VSMM),* 327–335.

McNamara, A., Chalmers, A., Troscianko, T., & Gilchrist, I. (2000). Comparing real and synthetic scenes using human judgements of lightness. *Proc. of EGWR 2000,* 207–219.

Meyer, G. W., Rushmeier, H. E., Cohen, M. F., & Greenberg, D. P. (1986). An experimental evaluation of computer graphics imagery. *ACM Transactions on Graphics, 5*(1), 30–35.

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

Smith, M. E. (1992). Neurophysiological manifestations of recollective memory experience during recognition memory judgements. *Journal of Cognitive Neuroscience, 5,* 1–13.

Stanney, K. M., Salvendy, G., Deisigner, J., DiZio, P., Ellis, S., Ellison, E., Fogleman, G., Gallimore, J., Hettinger, L., Kennedy, R., Lackner, J., Lawson, B., Maida, J., Mead A., Mon-Williams, M., Newman, D., Piantanida, T., Reeves, L., Riedel, O., Singer, M., Stoffregen, T., Wann, J., Welch, R., Wilson, J., & Witmer, B. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction, 10*(2), 135–187.

Travis, D. (1991). *Effective color displays.* London: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1–12.

———(1993). Varieties of consciousness and levels of awareness in memory. In A. D. Baddeley and L. Weiskrantz (Eds.), *Attention: Selection, awareness and control*—A tribute to Donald Broadbent, pp. 283–299. London: Oxford University Press.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments, 9*(5), 497–503.

Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and Virtual Environments, 7*(2), 129–143.

This article has been cited by:

1. Eric D. Ragan, Ajith Sowndararajan, Regis Kopper, Doug A. Bowman. 2010. The Effects of Higher Levels of Immersion on Procedure Memorization Performance and Implications for Educational Virtual EnvironmentsThe Effects of Higher Levels of Immersion on Procedure Memorization Performance and Implications for Educational Virtual Environments. *Presence: Teleoperators and Virtual Environments* **19**:6, 527-543. [Abstract] [PDF] [PDF Plus]
2. Beata Stahre, Monica Billger, Karin Fridell Anter. 2009. To Colour the Virtual World - Difficulties in Visualizing Spatial Colour Appearance in Virtual Environments. *International Journal of Architectural Computing* **7**:2, 289-308. [CrossRef]
3. Roy A. Ruddle, , Simon Lessels . 2006. Three Levels of Metric for Evaluating WayfindingThree Levels of Metric for Evaluating Wayfinding. *Presence: Teleoperators and Virtual Environments* **15**:6, 637-654. [Abstract] [PDF] [PDF Plus]
4. Doron Friedman, , Andrea Brogni , Christoph Guger , Angus Antley, , Anthony Steed , Mel Slater . 2006. Sharing and Analyzing Data from Presence ExperimentsSharing and Analyzing Data from Presence Experiments. *Presence: Teleoperators and Virtual Environments* **15**:5, 599-610. [Abstract] [PDF] [PDF Plus]
5. K. Mania, D. Wooldridge, M. Coxon, A. Robinson. 2006. The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics* **12**:3, 396-404. [CrossRef]
6. C.A. Zanbaka, B.C. Lok, S.V. Babu, A.C. Ulinski, L.F. Hodges. 2005. Comparison of Path Visualizations and Cognitive Measures Relative to Travel Technique in a Virtual Environment. *IEEE Transactions on Visualization and Computer Graphics* **11**:6, 694-705. [CrossRef]
7. Katerina Mania , Andrew Robinson , Karen R. Brandt . 2005. The Effect of Memory Schemas on Object Recognition in Virtual EnvironmentsThe Effect of Memory Schemas on Object Recognition in Virtual Environments. *Presence: Teleoperators and Virtual Environments* **14**:5, 606-615. [Abstract] [PDF] [PDF Plus]