Effects of visual cues and sustained attention on spatial presence in virtual environments based on spatial and object distinction

Sungkil Lee, Gerard Jounghyun Kim

Abstract

This article reports two human experiments to investigate the effects of visual cues and sustained attention on spatial presence over a period of prolonged exposure in virtual environments. Inspired by the two functional subsystems subserving spatial and object vision in the human brain, visual cues and sustained attention were each classified into spatial and object cues, and spatial and non-spatial attention, respectively. In the first experiment, the effects of visual cues on spatial presence were examined when subjects were exposed to virtual environments configured with combinations of spatial and object cues. It was found that both types of visual cues enhanced spatial presence with saturation over a period of prolonged exposure, but the contribution of spatial cues became more relevant with longer exposure time. In the second experiment, subjects were asked to carry out two tasks involving sustained spatial attention and sustained non-spatial attention. We observed that spatially directed attention improved spatial presence more than non-spatially directed attention did. Furthermore, spatial attention had a positive interaction with detailed object cues.

1. Introduction

1.1. General Introduction

Starting from the simple notion of “being there” (Heeter, 1992), presence has been developed into various multi-dimensional concepts over time. These concepts have been recently consolidated as “a perceptual illusion of non-mediation yielding a subjective sensation of being there in a mediated environment” (Alcainiz et al., 2002). This notion extensively covers broad and various concepts of presence such as spatial presence, social presence, and co-presence (Lombard and Ditton, 1997). Among such aspects of presence, this article focuses on the term, spatial presence, which is the sense of being in a physical space or virtual environment (VE) and having a mental representation of one’s own body as a part of a space or VE (Schubert et al., 1999; Regenbrecht and Schubert, 2002; Hofmann and Bubb, 2003).

Spatial presence in a VE inevitably requires media form factors and attentional allocation (Ijsselsteijn and Riva, 2003). The media form factors mediate physical experience relying on various technological cues for the external visual stimulation. They are largely modulated by attentional allocation (also called the “involvement” or “engagement”) that leads subjects to be sufficiently involved in the environment (Schubert et al., 1999; Lessiter et al., 2001; Ijsselsteijn and Riva, 2003). As regards the technological cues, a large body of literature has identified their contributions on presence, for instance: stereoscopy, motion, shadows, pictorial realism, 3D geometric detail, and 3D deformation (Slater et al., 1995; Hendrix and Barfield, 1996; Welch et al., 1996; Gerhard et al., 2001; Ijsselsteijn et al., 2001; Shim and Kim, 2003). These studies commonly support the argument that the more realistic and immersive the VE is, the greater the degree of presence is. Although sensorial realism induced solely from visual cues does not guarantee the greatest level of presence (Usoh et al., 2000; Heeter, 2003), we still anticipate that presence technology plays a pivotal role in content creation and delivery (Zhao, 2003). The greatest level of presence can be achieved from synergistic effects based on attractive contents and sufficient involvement.

However, most studies on presence constructs have only considered investigating their individual effects rather than their collective effects. Thus, an effective and economic design of a VE with regard to presence still remains difficult. For instance, given a limited amount of computational/hardware resources, it is unclear which presence elements to choose in order to induce relatively high spatial presence. Moreover, it is not clear how or if these presence elements interact with each other. Thus, studying presence constructs for their combination effects will contribute to both the design and evaluation of cost-effective VEs. This article aims
at discovering and understanding the collective effects of visual cues and attentional factors for designing VEs with high spatial presence.

In summary, our insights on visual cues and attention, as presence eliciting constructs, are simply that spatial presence is more dependent on spatial cues than non-spatial cues, and a user becomes more involved in the VE with spatially directed attention. These insights have their basis in the widely accepted neurological finding that the human brain has two functionally specialized subsystems subserving spatial localization and object recognition (Mishkin et al., 1983; Milner and Goodale, 1995; Deubel et al., 1998; Creem and Proffitt, 2001). Based on this functional separation, we classify the visual cues according to whether they are more involved in spatial localization or object identification. Likewise, attentional factors induced from tasks can be categorized into spatial and non-spatial. This distinction serves as a basis for our experiments and analysis.

1.2. Psychophysiological and theoretical background

This section presents related work, mainly psychophysiology, which is related to our hypotheses and experiments.

1.2.1. Spatial and object distinction of visual cues and attentional modulation

Various visual cues induce the sense of spatial presence in different ways. In order to deal with incoming visual stimuli, the human brain is functionally specialized for spatial and object vision. Spatial vision (often called the “where” system) mediates spatial localization for determining where an object is. Object vision (often called the “what” system) is responsible for object identification and recognition for determining what an object is (Mishkin et al., 1983; Deubel et al., 1998). Very recently, Baumbergarter et al. (2006) found, through an EEG study, that high spatial presence activated the parietal lobe regions in the brain, which mainly mediate spatial localization. This study supports the strong influence of spatial vision on spatial presence, which closely coincides with our insights.

The classic definition of the “where” and “what” systems anatomically maps them to the dorsal and ventral streams (Mishkin et al., 1983) in the cortical pathways of the human brain. The dorsal stream projects from the primary visual cortex (V1) to the posterior parietal cortex, whereas the ventral stream projects from the V1 to the inferotemporal cortex (IT). Numerous functional imaging studies have observed the cortical areas activated by specific visual cues that mediate object or spatial vision. For example, the binocular disparity and motion cues mainly evoke neural activities in the dorsal stream (DeAngelis and Newsome, 1999; Braddick et al., 2001), color, geometric shape, and shape deformation cues evoked activities in the ventral stream (Dubner and Zeki, 1971; Pasupathy and Connor, 2001; Kayser et al., 2005).

However, many researchers have recently argued that these anatomical streams are not precisely matched to the functional subsystems, and there exists a division of labor between the two streams (Milner and Goodale, 1995; Creem and Proffitt, 2001). Both streams process visual stimuli about spatial and object information but deal with them in different ways. The ventral stream transforms incoming visual information into perceptual representations of enduring characteristics of objects and their relationship, whereas the dorsal stream utilizes moment-to-moment information about objects to represent them within egocentric frames of reference (Milner and Goodale, 1995). However, in spite of interaction between the two streams, it is still clear that there exists a functional dissociation between the processing of spatial and object attributes, and they are more dependent on the dorsal and the ventral streams, respectively (Creem and Proffitt, 2001; Mazar et al., 2004).

Based on this functional distinction, we classify various visual cues according to whether they are involved in spatial localization or object identification. That is, if a given cue is more strongly involved in spatial localization, we classify it as a spatial cue. Conversely, if a cue is more strongly involved in object recognition, it is classified as an object cue. More specifically, the binocular disparity and motion perspective contribute to the spatial localization and depth perception in “personal” and “action” space (Cutting and Vishton, 1995). Thus, we categorize them as spatial cues. Cast shadows (shadows on a surface detached from an object) have been regarded as an important spatial cue in a pictorial representation (Yonas et al., 1978; Kersten et al., 1997). Thus, they are classified as spatial cues. Based on similar principles, geometric detail, texture, and shape deformation were categorized as object cues. The classification of visual cues utilized in typical VEs is given in Table 1.

Another important aspect is the effect of attentional modulation on either spatial or non-spatial visual information. Numerous early studies have found that distinct functional pathways (dorsal and ventral stream) and separate working memories are largely modulated in terms of selection-for-spatial-motor and selection-for-visual-perception controls (Mishkin et al., 1983; Treisman and Gormican, 1988; LaBerge and Brown, 1989; Posner and Petersen, 1990; Schneider, 1995; D’Esposito et al., 1998; Michie et al., 1999). Neuronal populations in either the dorsal or ventral stream have more activation for attentional highlights on either spatial or object attributes. These attentional highlights modulate the conjunction and speed of perception of visual stimuli (Treisman and Gelade, 1980; Vidyasagar and Parmer, 1999). Furthermore, these attentional controls are understood to be performed in a task-dependent manner (Deubel et al., 1998). This task dependency is consistent with the general statement that maintenance of a mediated illusion formed by bottom-up visual cues as well as top-down cognitive processes requires sufficient allocation of attentional resources to the computer-mediated VE (Witmer and Singer, 1998; Bystrum et al., 1999; Draper et al., 1999; Schubert et al., 1999; Regenbrecht and Schubert, 2002; Waterworth and Waterworth, 2001). This evidence has led us to differentiate attentional factors into spatial or non-spatial attention induced from the task or content factors in a virtual environment. That is, spatially directed attention boosts spatial perception (induced by spatial cues) and cognition capability. Accordingly, spatial presence becomes higher. On the other hand, with non-spatially directed attention, the sense of spatial presence would be mostly unaffected.

1.2.2. Spatial mental imagery and its relationship with spatial presence

As has been already alluded to, spatial and object cues/functions work in a different manner to form spatial presence, in other words, an adequate perception of oneself within physical space in an egocentric frame of reference. These different contributions can be naturally associated with the well-known environmental knowledge model for spatial mental imagery (Siegel and White, 1975; Thorndyke and Hayes, 1982; Ruddle et al., 1998; Waller et al., 1998). Analogous to the spatial and non-spatial distinction, environmental knowledge too can be separated into landmark and

<table>
<thead>
<tr>
<th>Cues</th>
<th>Categorization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereoscopics</td>
<td>Spatial</td>
</tr>
<tr>
<td>Binocular disparity</td>
<td>Spatial</td>
</tr>
<tr>
<td>Object motion</td>
<td>Spatial</td>
</tr>
<tr>
<td>Cast shadows</td>
<td>Spatial</td>
</tr>
<tr>
<td>Shape deformation</td>
<td>Object</td>
</tr>
<tr>
<td>Geometric detail (shape)</td>
<td>Object</td>
</tr>
<tr>
<td>Texture/Color</td>
<td>Object</td>
</tr>
</tbody>
</table>
spatial (or route and survey) knowledge (Epstein and Kanwisher, 1998). Also, each is differently distributed by functional subsystems on ventral and dorsal neural areas. Aguirre and D’Esposito (1997) demonstrated direct evidence of this correspondence between functional dissociation and landmark/spatial knowledge, by their fMRI experiments on “Position” and “Appearance” tasks.

Continuing to consider functional dissociation, we anticipate different roles of spatial and object cues during the formation of mental imagery and representation. The landmark knowledge mainly involves object recognition. Thus, object cues are used to identify the object and represent it in spatial working memory (Epstein and Kanwisher, 1998). In this stage, the user does not have a concrete sense of space, but only partial information. Thus, the spatial relationships within the surroundings are not adequately constructed, and accordingly spatial presence is relatively low. With continued exposure to the environment, these landmarks are connected together by spatial localization, and they are developed into spatial knowledge. Thus, the user obtains a concrete and greater sense of space and feels more immersed in the concrete virtual space.

Although this environmental knowledge model was proposed mainly for navigation in a large space, a recent PET study showed that this functional dissociation can be also applied to other types of spatial mental imagery such as mental exploration, construction, and scanning (Mazard et al., 2004). Thus, in the same context, we anticipate that spatial knowledge has a greater relationship to formation of a spatial mental model, and accordingly spatial cues are more relevant. However, the final perception of environments seems to be the result of the coherent interplay of object and space rather than disjointed aspects of them (Lee and Tversky, 2005). That is, we simultaneously perceive what an object is and where it is located. In this sense, object cues are not completely excluded, so they also aid in the formation of a spatial mental model. However, this occurs mainly at the initial stage in order to recognize the distinguishable appearance of an object (Mazard et al., 2004).

From these characteristics of spatial knowledge, we conjecture another notable characteristic of spatial cues: weak spatiotemporal learning capability. Spatial knowledge composed of neural representations has to be continuously and dynamically updated in conjunction with eye movements (Duhamel et al., 1992), whereas landmark knowledge is mainly mediated by location-invariant object cues.

Furthermore, illusory spatial information regarding the body position in VEs is typically different from space/motion predicted from real-world experience (Yardley, 1990). Thus, when a user experiences a great deal of sensory conflicts, these sensory conflicts often induce space/motion sickness (Oman, 1998). This is a severe side-effect of contemporary VR applications and can degrade presence as noted by Stanney et al. (1998). Fortunately, humans can adapt themselves and overcome sickness, given a sufficient amount of time and repeated exposure to the sensory rearrangements (Reason, 1978). Similarly to repeated spatiotemporal learning, this process also requires dynamic construction of a mental spatial model by suppressing one modality or fusing modalities together (Shimojo and Shams, 2001). This heavy and potentially conflicting sensory load makes learning of spatial knowledge much slower. Thus, we anticipate that spatial presence, even with rich spatial cues, would be rather low at the beginning of virtual experience, but become more relevant with prolonged exposure.

1.3. Hypothesis and general outline

Based on the distinction among visual cues and the weak perceptual learning capability of spatial cues, we hypothesize as follows: individual and synergistic effects from spatial and object cues enhance the level of spatial presence, but the effects of spatial cues are more relevant, particularly with longer exposure time. This hypothesis is tested in our first experiment. Subsequently, we investigate the effects of attention, in particular, sustained attention, which refers to the ability to maintain a fixed focus over extended periods of time. Sustained attention is a factor relevant to task-driven VEs, because the task primes the user to make deliberate detection of the target object/stimuli (Sarter et al., 2001). Our second hypothesis about attentional factors is that sustained attention which is spatially induced (e.g., through a spatial task) enhances spatial presence more than non-spatially induced attention. This hypothesis is tested in the second experiment.

Our hypotheses, if proven, can be used to conceive a practical strategy for VE construction so that a user is sufficiently immersed and can gain a concrete sense of place and experience. Strong spatial cues in VEs often require a high level of cost (both computationally and financially) and external devices. For example, a large and stereoscopic display (e.g., CAVE), animated motion, and shadows require expensive devices, careful calibration, and trained engineers. Authoring of attractive content or tasks to induce attentional involvement requires considerable efforts and expertise as well. Thus, we hope to contribute to establishing design guidelines for VEs which are useful for effectively integrating various types of spatial cues.

The rest of this article is laid out as follows. Section 2 presents our first experiment for investigating the effects of visual cues over a period of prolonged exposure. In Section 3, the second experiment on the effects of sustained attention is described. Finally, Section 4 discusses the general implication of our results and the conclusions of this paper.

2. Experiment I: effects of spatial and object cues over a period of prolonged exposure

In this section, we present a human experiment to investigate the effects of visual stimuli categorized into spatial and object cues over prolonged duration of exposure. Our hypothesis is that the spatial cues have stronger positive effects on spatial presence than the object cues with sufficient time of exposure. In addition to spatial presence, we measured the perceived size of the virtual space to assess the relationship between spatial perception and spatial presence.

2.1. Methods

2.1.1. Participants

Thirty-two paid graduate and undergraduate students (29 males and 3 females) participated in the experiment. Their ages ranged from 17 to 30, with an average age of 23.4 years. All participants had normal or corrected-to-normal vision, and they had prior experiences in VEs or 3D games. No sickness symptoms were observed during the experiment.

2.1.2. Apparatus and materials

The visual stimuli were presented to participants using a projection-based stereoscopic display (100-inch size and 1024 × 768 resolution) on the Intel Xenon 2.8 GHz processor and nVidia Quadro FX graphics card. Participants were seated 3 m away in front of the screen. CrystalEye shutter glasses (StereoGraphics, Inc.) were used as the stereoscopic imaging device, with a refresh rate of 120 Hz. All lights in the room were turned off during the experiment to minimize the distraction by real objects in the laboratory. The experimental devices and laboratory are shown in Fig. 1. A virtual underwater test environment was implemented using the WorldToolKit (Sense8, Inc).
2.1.3. Stimuli

Six visual cues were chosen for controlling the viewing conditions of the VE: (1) stereoscopy, (2) object motion, (3) cast (detached) shadows, (4) shape deformation of objects, (5) texture, and (6) 3D geometric detail. Among the cues, the first three were classified as spatial cues (S), and the others as object cues (O). Each cue had two levels. The High S (HS) condition involved stereoscopy, cast shadows, and fast objects motion. On the other hand, for the Low S (LS) condition, monoscopy without cast shadows and minimal object motion were used. The High O (HO) condition was configured with object deformation, textured surfaces, and high geometric detail, and vice versa for the Low O (LO) condition. Table 2 and Fig. 2 show the details and examples of the experimental conditions. With the two levels for each factor, \( \frac{2^2}{C_2^2} \) combinations

![Fig. 1. Experimental devices and laboratory used in Experiment 1. The picture was brightly taken for illustration purpose, whereas the actual experiment was carried out under dark lighting in order to minimize distraction by other objects in the laboratory.](image1)

![Fig. 2. Virtual underwater as the test environment in Experiment 1. Each image represents the corresponding experimental condition. Note that the images are not able to illustrate the stereoscopy and motion cues included in HS condition.](image2)

<table>
<thead>
<tr>
<th>Condition</th>
<th>S (spatial cues)</th>
<th>Motion (fish)</th>
<th>Cast shadows</th>
<th>O (object cues)</th>
<th>Shape deformation</th>
<th>Geometry (no. polygon)</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSLO</td>
<td>No</td>
<td>Stays in place</td>
<td>No</td>
<td>No</td>
<td>Low</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HSLO</td>
<td>Yes</td>
<td>Moves around</td>
<td>Yes</td>
<td>No</td>
<td>Low</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>LSHO</td>
<td>No</td>
<td>Stays in place</td>
<td>No</td>
<td>Tail wagging</td>
<td>High</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HSHO</td>
<td>Yes</td>
<td>Moves around</td>
<td>Yes</td>
<td>Tail wagging</td>
<td>High</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Stereoscopy, object motion, and cast shadows belong to spatial cues, and shape deformation, geometric detail, and texture belong to object cues.
were tested in the experiment, that is, LSLO, LSHO, HSLO, and HSHO. The presentation order of the four sessions was balanced using Latin squares.

2.1.4. Design

The experiment used a 2 (spatial cues) × 2 (object cues) × 10 (duration of exposure) within-subject repeated-measure design. Duration of exposure (E) was also manipulated such that subjects were asked to rate the subjective level of spatial presence (SP) and perceived size (PS) of the virtual space every minute during a 10-min session. Ten minutes were chosen based on pilot experiments that established that after five or more minutes, spatial presence for some participants decreased, whereas for the others, it slowly increased or became saturated. Many participants belonging to the former case indicated that this was due to boredom, that is, only looking at a fixed place without any task. Even the latter participants showed a decreasing trend after 10 min of exposure. Thus, considering these results, we chose 10 min as the maximal duration of exposure.

2.1.5. Measurement

We had to measure the dependent variables with minimal delay during run-time, and therefore it was impossible to use a comprehensive presence questionnaire such as one by Witmer and Singer (1998) and Usoh et al. (2000). Thus, we used one direct question about spatial presence: “How much do you have a feeling of being in underwater? Rate your feeling on a 100-point scale.” For the same reason, the perceived size was assessed in the same way, “Rate the perceived size of the VE on a 100-point scale.” For rating the perceived size, a 50 on a 100-point scale was referenced as representing the “same” size as the corresponding object in the real world. For example, a 25-cm real fish was used as a reference for the size perception test of the virtual fishes. Thus, if the virtual fish was felt larger than, same as, smaller than the reference, the solicited score would be greater than 50, equal to 50, or less than 50, respectively. We emphasized that the subject should not belatedly adjust the score and just rate as he/she felt at the moment, when responding to the question.

In order to measure spatial presence continuously, we asked the participants for verbal ratings every minute (recorded by the experimenter), similar to the methods used by Garau et al. (2004). Although this method sometimes results in a break-in-presence (Slater and Steed, 2000), we considered this negative effect to be less than that of a continuous measure such as the slide bar method (Ijsselsteijn and de Ridder, 1998). The slide bar method allows better observation of an immediate fluctuation in presence, but it places sustained mental load on the user that can have a negative effect on spatial presence itself, due to attention being diverted to a non-spatial attribute (i.e., the score of spatial presence). Furthermore, Garau et al. (2004) observed that these kinds of break-in-presence were rapidly recovered from, in many cases, a few seconds after the verbal answer.

2.1.6. Procedure

On arriving at the laboratory, a participant first received instructions to carefully watch the virtual underwater environment and orally rate the levels of spatial presence and perceived size of the virtual space every minute for 10 min. After the instructions were given, the participant was asked to sit in front of the projection screen and to wear shutter glasses. To balance the luminance differences between LS and HS conditions, the subject wore shutter glasses even under the monoscopic display condition. After this, the first session started. A text signal asking for a rating appeared at the bottom of the screen once every minute. The subject’s answers for the two questions were recorded by the experimenter. After finishing the session, the participant took a rest, and the same procedure was repeated for the remaining three sessions.

2.1.7. Nonlinear regression with von bertalanffy growth function for growth patterns

During analysis of the experimental data, some growth and saturation patterns were found along with a duration of exposure factor. In order to analyze these patterns, we applied nonlinear regression with the von Bertalanffy growth function (VBGF), which is widely used for describing an asymptotic growth and saturation pattern (von Bertalanffy, 1938). The original form of VBGFs is defined as:

\[ y = L_\infty - (L_\infty - L_0) e^{-Kx}, \]

where \( L_\infty, L_0, \) and \( K \) represent an asymptotic limit, initial intercept at \( x = 0 \), and rate constant, respectively. This function is very effective for analyzing a growth pattern and finding a saturation point in time.

Ideally, asymptotic growth data become fully saturated with an infinite duration of exposure. Therefore, we needed to define reasonably saturated points. In our analysis, 95% of \( L_\infty \) was used as the threshold for determining saturation. Under 95% threshold, we let \( y \) be 0.95\( L_\infty \). Then, a saturation point, \( x_{\text{threshold}} \), is determined as:

\[ x_{\text{threshold}} = \frac{1}{K} \ln \left( \frac{L_\infty - L_0}{0.05 L_\infty} \right). \]

2.2. Results

2.2.1. Spatial presence (SP)

The measured ratings of SP for three factors—S, O, and E—are summarized in Fig. 3. The mean ratings for each level are shown along with the standard errors. Overall, the mean SP with the high level(s) increased for all factors. The HS and HO conditions resulted in significantly higher ratings than LS and LO, respectively. Also, the SP increased and gradually saturated with increasing E. In order to understand the statistical significance of these effects, we applied repeated-measures ANOVA, and the results are summarized in Table 3. Inter-trial dependencies caused by the repeated measures were removed using the SAS Proc Mixed (SAS Institute Inc., 1990). All the main effects of S, O, and E were statistically significant, and a significant interaction was observed only between S and E.

As has been already mentioned, we observed a saturating trend for the SP with the increasing E. To examine this effect in more detail, we applied nonlinear regression between E and SP using the VBGF and obtained their relationship as:

\[ \text{SP} = 54.8 - (54.8 - 39.8) e^{-0.01E}, \]

where 54.8, 39.8, and 0.014 corresponded to \( L_\infty, L_0, \) and \( K \), respectively, and \( R^2 = 0.9983 \). Based on this formulation, we establish that the overall saturation of the SP has occurred at 6.97 min (see the vertically dotted line in Fig. 3).

An interaction effect was observed between S and E. The implication is that the effects of E differ with the levels of S cues. In order to examine the individual effects of LS and HS, we graphically compared the effects of the increasing E. Fig. 4 shows the result. We can intuitively grasp that the SP increased more steeply under HS than under LS in the initial stage of exposure (less than 4 min), and the difference between the two trends was maintained afterwards. SP ratings under HS and LS were also regressed with the VBGF, and growth patterns \( (R^2 = 0.9947 \text{ and } 0.9894 \text{ under HS and LS, respectively}) \) were attained. The estimated coefficients are presented in Fig. 4. Saturation points were similarly found at 5.2 (a vertical line in Fig. 4) and 13.5 min (not shown in the figure), respectively. For
the two regressed VBGFs, \( K \) values (0.32 and 0.14 under HS and LS, respectively) also confirm these trends (a VBGF with a higher \( K \) value increases more steeply).

To further assess learning behavior and the different roles of S and O during the formation of spatial presence, we carried out a correlation analysis similar to the method of Ackerman (1988). For each fixed duration of exposure, we separately calculated the Pearson correlation coefficients of S and O for SP. Fig. 5 effectively illustrates the changing contributions of S and O in terms of the correlations with SP. The figure shows that the correlation coefficients for O remained somewhat constant or marginally decreased. However, the coefficient for S steeply increased up until about 4 min, then it marginally decreased and remained roughly con-

![Fig. 3](image1)

**Fig. 3.** Mean ratings of spatial presence (SP) for each level of S, O, and E factors. The vertical error bars indicate standard errors. In (c), the curve was regressed with von Bertalanffy growth function (VBGF), and the vertical dotted line at 6.97 min represents the saturation point under the 95% threshold.

![Fig. 4](image2)

**Fig. 4.** Mean ratings of SP under LS and HS conditions along the levels of E. Ten points in each condition were regressed with VBGF. Under the 95% threshold, the saturation points for LS and HS were found and marked as a dotted line at \( x = 13.5 \) (not drawn in the graph) and \( x = 5.2 \) min, respectively.

![Fig. 5](image3)

**Fig. 5.** Scatter plot of Pearson correlation coefficients between S and O for SP in Experiment I. Each coefficient was fitted with a cubic polynomial. The upper curve represents the fitted polynomial for S and SP \( (R^2 = 0.8924) \), and the lower curve, for O and SP \( (R^2 = 0.5916) \).

Table 3

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>31</td>
<td>214.15***</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>108.80***</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>E</td>
<td>9</td>
<td>40.45***</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>S × O</td>
<td>1</td>
<td>0.44</td>
<td>.5106</td>
</tr>
<tr>
<td>S × E</td>
<td>9</td>
<td>1.89</td>
<td>.0491</td>
</tr>
<tr>
<td>O × E</td>
<td>9</td>
<td>0.10</td>
<td>.9996</td>
</tr>
<tr>
<td>S × O × E</td>
<td>9</td>
<td>1.40</td>
<td>.1812</td>
</tr>
</tbody>
</table>

S, O, and E represent spatial and object cues, and duration of exposure, respectively. Note. *p < .05, **p < .01, ***p < .001. The data were analyzed using the SAS Proc Mixed (SAS Institute Inc., 1990), which does not yield sum of squares.
stant, but at higher levels than those of O. This clearly shows that the contribution of S on the SP is higher than O. Also, the effects of S did increase, especially in the initial stage, whereas O had a constant influence on spatial presence during the entire session.

2.2.2. Perceived size (PS) of the virtual environment

The perceived size of the virtual space was analyzed in a manner similar to spatial presence. In summary, PS also increased with longer exposure. However there were irregular fluctuation and no trend of saturation. Thus, no explicit relationships between visual cues and duration of exposure were found. As one of the aims was to investigate the relationship between perceived spatial size and spatial presence, we applied linear regression between the two (see Fig. 6). A statistically significant linear regression coefficient was observed between the two, explicitly showing a positive correlation value. This result is consistent with the findings of Hofmann et al. (2001); relative size perception increases with higher spatial presence.

2.3. Discussion

Our results indicated that spatial presence was closely related to both spatial and object cues. As expected, spatial presence increased with the provision of both rich spatial and object cues. In addition, no interactions between the two types of cues were found, which supports our initial premise that spatial and object cues mediate spatial presence in different ways, similar to their functional division in the brain. Moreover, this is also consistent with the model of spatial mental imagery described in Section 1.2. Both spatial and object cues contribute to the formation of egocentric space model, but their functions are differentiated: object cues are mainly related to landmark knowledge, whereas spatial cues are related to spatial knowledge.

We further found that spatial presence increased more with the existence of spatial cues than with object cues, which coincided with our initial hypothesis. The difference in the means (Fig. 3) and our correlation analysis also supported this finding. This difference in the relative contributions can be explained by the functional distinction of visual cues (Mishkin et al., 1983; Baumgartner et al., 2006). That is, spatial cues are more closely linked to the formation of an egocentric mental model. Thus, with provision of rich spatial cues, the spatial mental model becomes more intense and concrete. Accordingly, the feeling of spatial presence is enhanced. We note that the exact degree of increase in spatial presence cannot be determined (by the existence of spatial or object cues), because enforcing the same degrees of scaling for the different test conditions was not possible. These relative effects can be varied for specific configurations by controlling various details or existences of the visual cues.

Regarding the duration of exposure, we found that spatial presence increased and saturated with longer exposure time. Satturation points can be varied for specific configurations, but we believe that this general trend is valid. Furthermore, we observed that the relative contribution to spatial presence of the spatial cues increased more steeply in the initial stage of exposure than that of the object cues. In the latter stage, after a sufficient passage of time, the contributions remained roughly constant, as was the case for the object cues.

This behavior is effectively explained by the fact that humans have weak spatiotemporal learning capability. At the initial stage of an immersive experience, the currently limited technology generates a great deal of sensory/spatial mismatch (in particular, from the artificial stereoscopy), and thus the subject needs time to adapt and properly update the mental space model. This update process is not immediate, and is in fact much slower in a VE than in the real world (Ruddle, 2001). In any case, spatial cues are more important than object cues in this process. As the model of mental space is updated and spatial mismatches are resolved with repeated and prolonged exposure (Reason, 1978; Stanney et al., 1998; Smeets et al., 2002), a user will become more aware of his/her spatial location in the virtual space. Accordingly, spatial presence becomes higher. After some time when the mental space model is more or less complete, this dynamic updating slows down and the distinction among different types of cues becomes irrelevant.

On the other hand, during the dynamic updating of mental space, object cues are mainly involved in the identification of objects. Landmark knowledge is likely to acquire structural saliency by distinguishing features from other landmarks (Mazzard et al., 2004; Klippel and Winter, 2005). Thus, it depends on identification of a rough object shape rather than accurate memorization of a detailed appearance. Hence, appearance coding by object cues is an immediate process relying less on temporal adaptation, and accordingly, its contribution on spatial presence stays rather constant during the entire process.

The correlation analysis revealed a slight tendency for a decreasing contribution from cues (mainly for object cues). We believe that this could be a side effect of boredom. That is, our experiment lacked an explicit task, and thus a user was likely to become bored (and they also reported that this was the case) due to the relatively long duration of the experiment time.

To summarize, our findings suggest that spatial presence is related to spatial and object cues as follows:

1. both the spatial and object cues increase spatial presence and their main effects are additive,
2. the effect of the spatial cues is more relevant than that of object cues,
3. the contribution of spatial cues is dependent on exposure duration, whereas that of object cues is not, and
4. the contribution of spatial cues steeply increases in the initial stage but remains roughly constant afterwards.

3. Experiment II: effects of sustained attention

This section presents our second experiment to investigate the effects of sustained attention induced from task factors on spatial presence. This experiment aims to further relate or take advantage of spatial cues to enhance spatial presence through a spatially attended task. We apply the distinction between spatial and non-spatial attention, analogous to the distinction between visual cues.

![Fig. 6. A scatter plot and regressed line of spatial presence (SP) versus perceived size (PS) for 40 (2 x 2 x 10) data points. We can intuitively grasp that the higher the SP is, the higher the PS. The regressed line reveals a statistically significant increase of PS by SP (B = 0.36, SE B = 0.02, t = 9.94, t = 16.66, p < .0001, and R² = 0.8796).](image)
Our hypothesis is that spatial attention has greater positive effects than non-spatial attention on spatial presence. The distinction between visual cues used in the first experiment is also adopted, and possible interactions between the effects are investigated as well.

3.1. Methods

3.1.1. Participants and materials

Twenty-four paid students (19 males and 5 females) participated in the experiment. Their ages ranged from 19 to 28, with an average of 23.25 years. All the participants had prior experiences in VEs or 3D games. No sickness symptoms were visible during the experiment. The same platform as that in Experiment I was used, but, instead of the virtual underwater environment, a virtual office was used as the test environment (see Fig. 7).

3.1.2. Stimuli

The distinction between spatial and object cues was applied again to establish if there was any interaction between the effects of visual cues and attentional factors. Among the six visual cues used in Experiment I, we used only four cues: (1) stereoscopy, (2) shadows, (3) texture, and (4) geometric detail. The level of cues was similarly controlled, but in the case of texture, low-resolution textures were used under the LO condition instead of a single color. Object motion and deformation were excluded from the experiment, because they were not required for the given tasks.

In a manner similar to Experiment I, the effects of user interaction were also precluded. Instead of the fixed viewpoint in Experiment I, the viewing direction was automatically controlled via passive navigation along a predefined path. The navigation cycle was repeated about seven times. In order to measure spatial presence in a reasonably saturated state, one session was performed for 10 min based on the result of Experiment I (estimated saturation time was 6.97 min). In Experiment I, the LO condition reached saturation around at about the 13 min mark. However, the 10 min mark is still within the limits of the 92% threshold, which can be regarded as reasonably saturated. Participants were first immersed for 5 min, and then one of the two tasks was presented in a balanced order.

Non-spatial and spatial sustained attention were evoked by employing two types of tasks, (1) counting the number of pens in the VE (non-spatial task; NT) and (2) memorizing and comparing the spatial locations of objects in the VE (spatial task; ST). Both tasks were designed to engage the user in heavy sustained attention during the entire session. The counting task mainly required identification of colors and numbers of pens, and their spatial locations were mostly irrelevant to the required visual search. Thus, the task was primarily non-spatial in nature. For the spatial task, the participants had to memorize the spatial locations of objects. Thus, spatial localization of objects relative to the observer was very important in the task. The number of juxtaposed objects (e.g., 32 objects) was sufficient to maintain a heavy mental workload throughout the task.

Fig. 8 shows an example of the non-spatial task under the LSLO condition. After the first 5 min, a text signal appeared on the screen for 5 s. Then, four colored pencils appeared in the virtual office and subjects were asked to separately count the numbers of pencils for the two specified colors. The saliency and contrast between the two colors and numbers of the pens
cils under each condition were set as follows: 9 green and 12 red under LSLO, 6 blue and 8 green under LSHO, 10 red and 10 yellow under HSLO, and 10 blue and 11 yellow under HSHO.

An example of the spatial task under the LSLO condition is shown in Fig. 9. For the first 5 min, subjects were asked to observe the virtual office and to memorize the spatial locations of objects. After a text signal appeared, the spatial locations and existences of the 32 objects were randomly changed, and the viewpoint was reset to the starting point. During the remaining 5 min, the subjects were asked to find and memorize differences in the newly juxtaposed environment.

3.1.3. Design

The experiment used a $2 \times 2 \times 2$ within-subject design. The independent variables were: (1) type of sustained attention (T), (2) level of spatial cues (S), and (3) level of object cues (O). The dependent variable was also the subjective level of spatial presence (SP), i.e., the degree of feeling of being in the virtual office. The presentation order of the eight sessions was also balanced using Latin squares.

3.1.4. Measurement

Spatial presence was measured in a similar manner to Experiment I. Table 4 shows the questionnaire used. Depending on the given task, Question 1 asked about the number of pencils corresponding to the specified colors or spatial differences in the juxtaposed VE. Question 2 directly asked about the level of spatial presence experienced by the user on a 100-point scale. In order to quantify the task performance, the counting error was tabulated for the non-spatial task. The number of user-detected spatial differences in locations and existences of the objects was tabulated for the spatial task.

3.1.5. Procedure

Having received instructions for the experiment, a participant was seated and asked to carefully watch the screen in a fixed viewing position with no user interaction. The subject was asked to watch the virtual office for the first 5 min. That is, the subject was immersed in order to adapt to the VE. At the beginning of the sixth minute, a text signal appeared on the screen requesting that a task be performed, and the subject carried out the task for the remaining 5 min. After finishing the session, the subject was asked to rate the level of spatial presence on the printed questionnaire and provide answers for the task (the numbers of colored pencils or differences in the spatial configurations).

3.2. Results

3.2.1. Spatial presence

The measured ratings of SP with respect to the three factors—S, O, and T—are summarized in Fig. 10. The mean ratings for each level are shown along with the standard error bars. Overall, the inclusion of high level cues increased SP for all three factors. Under the ST condition, SP was marginally increased (approximately 2.4). Both the HS and HO conditions resulted in a similar increase in SP, as in the results of Experiment I. In order to examine the relative contributions of the three factors, $2 \times 2 \times 2$ ANOVA was applied to the data, and the result is shown in Table 5. As expected, both S and O showed very significant differences ($p < .0001$) between the corresponding two levels. Also, T demonstrated a statistically significant increase between NT and ST, which means that the spatially directed task improved the sense of spatial presence.

An interesting interaction was observed between O and T. To investigate further, we compared the differences in a post-hoc analysis. For each condition of O, we separately plotted the differences between NT and ST (see Fig. 11). The figure shows a significant increase of SP with ST under the HO condition (about 4 over the 100-point scale). We further performed a simple main effect test to confirm the statistical significance between NT and ST under the HO condition by one-way ANOVA, $F(1,23) = 11.44, p = 0.0026$. This interaction effect implies that spatial attention improves spatial presence if a high level of object cues are provided.

3.2.2. Task performance

For each task, two different metrics were applied to evaluate performance. For NT, the absolute difference between the correct number of pencils and the user-counted number was used (i.e., the greater the difference, the lower the performance). For ST, the number of correctly identified juxtaposed objects was used as the task performance.

Tables 6 and 7 show the mean error estimates and the ANOVA source table for NT conditions, respectively. Only O resulted in a significant effect on the task performance. In fact, the lower the level of object cues was, the better the performance (i.e., lower error). This can be attributed to the fact that the counting task was easier with simple looking objects than with realistic objects. This indirectly indicates that spatial presence is not correlated with the performance of the non-spatial task. Our additional correlation analysis between the two supports this argument ($r = 0.0392, p = 0.70$).

---

Table 4

<table>
<thead>
<tr>
<th>No.</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>How many green and red pencils were in the office? (non-spatial task)</td>
</tr>
<tr>
<td>1.2</td>
<td>What was the spatial difference (existence, location, or orientation) after the environment was juxtaposed? (spatial task)</td>
</tr>
<tr>
<td>2</td>
<td>How much did you feel that you were being in the virtual office? (0–100)</td>
</tr>
</tbody>
</table>

The colors of pencils in Question 1.1 varied with the conditions. Under LSLO, LSHO, HSLO, and HSHO conditions, green/red, blue/green, red/yellow, and blue/yellow pairs were used, respectively. Note: Question 1 alternated 1.1 and 1.2 according to the kind of the task.

---

Fig. 9. An example of the spatial task under the LSLO condition. A subject was immersed and asked to memorize the locations of objects for the first 5 min, and then a text signal appeared for 5 s. During the remaining 5 min, the subject was asked to find any changed objects in the juxtaposed environment; in this example, a chair, a desk, green paper boxes, a beverage can, a copier, an air-conditioner, a printer, a monitor, a keyboard, and a cabinet were changed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)
Tables 8 and 9 show the mean accuracies and ANOVA source table under ST conditions. Under ST, the performance of subject groups in the extreme ends of the spectrum (i.e., HSHO and LSLO conditions) was significantly higher than that observed in the other two. It appears that when spatial and object cues were congruent, subjects performed better, but when spatial and object cues were not, this caused a decline in performance.

3.3. Discussion

Our results indicate that spatially directed attention improves spatial presence more than non-spatial attention does, which is consistent with our expectation. Moreover, spatially attentive task or contents can improve spatial presence. This result establishes that synergistic effects between technological spatial cues (e.g., visual spatial cues such as stereo, shadow, and motion) and an attentional task can occur.

Furthermore, the positive interaction between spatial attention and object cues is a very interesting result. This implies that, if the appearance of things in VEs are presented in great detail (object cues), spatial presence can be improved by spatial attention. In other words, spatial attention induced by consciously designed content could compensate for insufficient spatial cues, e.g., even in a conventional desktop system (with no fancy VR device support). This implication is consistent with other recent findings that spatially directed content (e.g., roller coaster simulation) improves spatial presence without expensive spatial cues such as stereoscopy and a large screen (Baumgartner et al., 2006). This is very useful information for VE designers who are faced with the question of whether to provide expensive high spatial detail, while a non-spatially rich environment can be achieved with relatively less efforts. For instance, modeling of detailed geometry with high-quality textures is less costly in practice compared to stereoscopic imaging in

Table 5
Source table of within-subject 2 × 2 ANOVA for S, O, and sustained attention (T)

<table>
<thead>
<tr>
<th>Sources</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>23</td>
<td>44925.75</td>
<td>1953.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>14752.55</td>
<td>14752.55</td>
<td>42.26 ***</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>O</td>
<td>1</td>
<td>20315.75</td>
<td>20315.75</td>
<td>83.71 ***</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>263.67</td>
<td>263.67</td>
<td>7.92</td>
<td>.0098</td>
</tr>
<tr>
<td>S × O</td>
<td>1</td>
<td>5.67</td>
<td>5.67</td>
<td>0.06</td>
<td>.8157</td>
</tr>
<tr>
<td>S × T</td>
<td>1</td>
<td>43.13</td>
<td>43.13</td>
<td>0.97</td>
<td>.3347</td>
</tr>
<tr>
<td>O × T</td>
<td>1</td>
<td>128.38</td>
<td>128.38</td>
<td>4.94</td>
<td>.0364</td>
</tr>
<tr>
<td>S × O × T</td>
<td>1</td>
<td>39.42</td>
<td>39.42</td>
<td>0.97</td>
<td>.3349</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.

Fig. 10. Means and standard errors of SP for each level of the S, O, and sustained attention (T) in Experiment II.

Fig. 11. Differences of SP under the NT and ST conditions, with standard error bars, as a function of LO and HO in Experiment II. Under the HO condition, means of SP significantly differ between the NT and ST conditions.

Table 6
Mean error estimates as a function of S and O for non-spatial task (NT)

<table>
<thead>
<tr>
<th>O</th>
<th>S</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3.04</td>
<td>2.10</td>
<td>2.17</td>
</tr>
<tr>
<td>High</td>
<td>3.67</td>
<td>2.04</td>
<td>3.79</td>
</tr>
</tbody>
</table>

A higher mean error indicates a lower accuracy.

Table 7
ANOVA source table for non-spatial task performance

<table>
<thead>
<tr>
<th>Sources</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>23</td>
<td>150.83</td>
<td>6.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>3.38</td>
<td>3.38</td>
<td>0.77</td>
<td>.3900</td>
</tr>
<tr>
<td>O</td>
<td>1</td>
<td>30.38</td>
<td>30.38</td>
<td>13.94 **</td>
<td>.0011</td>
</tr>
<tr>
<td>S × O</td>
<td>1</td>
<td>6.00</td>
<td>6.00</td>
<td>2.58</td>
<td>.1219</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.
a large display, rendering of multiple screens, and expensive computation for high-quality animation and shadows.

Regarding the effects of the visual cues, similar results to Experiment I were obtained. That is, inclusion of rich spatial and object cues increased spatial presence again, although the motion cue was not included. One notable result is that the increase caused by spatial cues was reduced compared to that of Experiment I. This might be due to the passive navigation, which was required for the visual search task in the experiment. This passive navigation could have induced some degree of ego-motion (vection) due to relative motion perception in the observer’s view (von der Heyde and Riecke, 2002). In this sense, the difference between spatial and object cues was somewhat reduced. Additionally, we note that, if object cues do not maintain consistent details (e.g., due to changing viewpoint, motion blur), this may lead to a loss of feeling of presence.

The results on the performance of the non-spatial and spatial task indicate that the two measures have no causal relationship with spatial presence. Non-spatial task was negatively correlated to the fidelity of object cues (i.e., higher fidelity was a distraction to the non-spatial task), but no correlation was found to spatial presence. As for the spatial task, the important factor was the congruence between the object and spatial cues rather than the fidelity of the individual cues or spatial presence. These findings suggest that efforts to increase spatial presence do not necessarily improve the task performance, and it would be productive to manipulate cues directly related to the task performance as also noted by Welch (1999).

4. Conclusions and general discussion

In-depth understanding of spatial presence can serve as a basis for designing and evaluating VR applications and even for traditional desktop applications, including 3D geometric modelers, games, and other human–computer interfaces. In this article, based on the functional dissociation subserving spatial localization and object recognition in the human brain, the visual cues and sustained attention were each classified into spatial and object cues, and spatial and non-spatial attention, respectively. According to the distinction, we have investigated the effects of visual cues and sustained attention on spatial presence over a period of prolonged exposure. The first experiment found that spatial cues were more relevant than non-spatial cues for evoking a feeling of spatial presence. Furthermore, spatial cues made a greater contribution to spatial presence after sufficient saturation. The second experiment about sustained attention revealed that spatial attention enhances spatial presence more than non-spatial attention. Furthermore, it has a positive interaction with rich non-spatial cues.

With respect to spatial presence, spatial and non-spatial distinction for either technological cues or attention has a special distinction for either technological cues or attention has a special significance. The former can strive to improve a VR system by providing sufficient spatial cues. Also, the content artists can still save a poorly designed VR system by the production of spatially directed content. In these respects, our findings could serve as good guidelines for designing VR systems.

Growth and saturation phenomenon of spatial presence has an important implication for presence measurement. A number of studies have attempted to develop devices for measuring the level of presence, but many of them have overlooked its saturation aspects. Presence measurement, either subjective or objective, will reveal its true effects and any interactions between pertinent factors, only after the user is saturated and fully immersed in the mediated environment after sufficient exposure.

In the future, we plan to extend our work, relating human spatial ability to spatial presence to broaden our understanding. Also, interaction factors, overlooked in our study, will be considered for their influence on the sense of spatial presence with a varied combination of visual and attentive cues.

References


Ijsselsteijn, W., de Ridder, H., Freeman, J., Avons, S.E., Bouwhuis, D., 2001. Effects of stereoscopic presentation, image motion, and screen size on subjective and


