16 Presence in Industrial Virtual Environment Applications — Susceptibility and Measurement Reliability

Jan HOFMANN, Heiner BUBB

Abstract In this chapter, we report the concept and results of an empirical study that explored the effects of immersion and pictorial realism on the sense of presence in a virtual environment (VE). The VE used was similar to a typical industrial application. Seventy-seven participants each experienced one of four settings with varied combinations of immersion and pictorial realism, displayed in a projection based virtual reality system. Their sense of presence was measured using a post-exposure questionnaire developed by Schubert et al. [1]. It allows to differentiate between three presence facets — reality appraisal, involvement, and spatial presence. A repeated measurement enabled us to study the reliability of our presence data. Concerning the susceptibility of presence, our results show significant effects of both immersion and pictorial realism on two of the presence facets. Furthermore, the influence of immersion proved to be dependent on the degree of pictorial realism displayed. Practical implications of these results for the design of virtual reality systems are proposed. Finally, we point to some implications of our findings for the conceptualization of the sense of presence in virtual environments.

Contents

16.1 Motivation of this study ................................................................. 238
16.2 Methodology — experimentation in an industrial setup ........... 238
16.3 Presence measurement reliability — capturing a volatile state 241
16.4 Susceptibility of presence — on the need to differentiate ........ 243
16.5 Some implications for the conceptualization of presence ....... 245
16.6 Conclusions .................................................................................. 245
16.7 References .................................................................................. 246
16.1 Motivation of this study

In recent years, the sense of presence in virtual environments (VEs) has gained much interest in various research communities. Nevertheless, (a) the factors affecting this cognitive state, (b) the methods of measuring it, (c) its effects on other parameters, as well as (d) its very nature are still controversial issues. In this chapter, we take a closer look at aspects (a), (b), and (d). To this end we report results of experiments we conducted in a VE that is similar to an industrial application: a virtual car cockpit.

It has been suggested — and in some cases shown experimentally — that the degree of presence can be influenced by diverse factors. The objective of our experiments was to analyze interactions of effects of immersion and pictorial realism (PR) on the sense of presence. The term immersion is used here — as e.g. by Slater and Wilbur [3] — to describe characteristics of the virtual reality system, of the virtual environment, and other factors that enhance the sense of presence. Pictorial realism is usually regarded to be one of these factors (see e.g. [4]). In this study, we varied pictorial realism separately from other immersion factors and thus excluded it from our definition of immersion.

This kind of analysis is particularly useful from a practical point of view. Many researchers suppose effects of presence on the user’s behavior in the VE, e.g. on his or her task performance [5]. But in industrial VE applications, both immersion and pictorial realism can vary widely. What happens to the users’ sense of presence if immersion is increased, but pictorial realism is low, or vice versa? Understanding this would be valuable for designers and operators of VE applications with certain demands regarding the sense of presence the users shall experience. Furthermore, immersion and pictorial realism cause different types of costs in their generation.

We thus designed the experimental settings of this study to reflect ‘archetypes’ of industrial VEs, providing different combinations of immersion and pictorial realism. This enables us to map archetype settings and resulting degrees of presence, which in turn can facilitate effective allocation of resources.

Three crucial questions remain to be answered, though: Which kind of ‘presence’ are we talking about, what is the most appropriate way of measuring it, and, finally, how reliable are our measurements? The former two questions are closely interlinked and can be answered with respect to the objective and constraints of our experiments. Determining the reliability of the kind of measurement used proves to be somewhat tricky, though. We therefore devote an own section to this topic before we discuss the effects of immersion and pictorial realism on our presence measurements.

Having considered possible explanations for our empirical findings, we eventually point to some implications our results might have for the conceptualization of the sense of presence.

16.2 Methodology — experimentation in an industrial setup

16.2.1 Varying immersion and pictorial realism

To generate low and high values of immersion and pictorial realism, we varied several factors simultaneously. The choice of these factors followed a compromise strategy: (a) We looked for factors whose variation was expected to exert a strong impact on the sense of

---

1 Parts of the work reported in this chapter have originally been presented at Human Computer Interaction International 2001, New Orleans [2]. The study was conducted by the Society and Technology Research Group, DaimlerChrysler AG, in co-operation with Passenger Car Development Mercedes-Benz, DaimlerChrysler AG, and the Chair of Ergonomics, Technische Universität München.
presence. Previous results of diverse theoretical and experimental studies were considered to try and predict their impact (e.g. [3, 4, 6-10]). (b) Factors that would not be varied in a practical application of a system like ours were not considered. E.g., head tracking would not be switched off deliberately, though this might have a profound effect on the sense of presence. Table 16.1 summarizes the factors chosen and their respective ‘low’ and ‘high’ values (see also Figure 16.1).

The four different ‘high’ / ‘low’ combinations of immersion and pictorial realism resulted in the four experimental settings. The virtual scene used in all four settings was the front half of a passenger car interior actually in production. It was based on the original data used in the product development process and was displayed in 1:1 scale. The virtual interior was combined with a real driver’s seat and steering wheel to enhance participants’ impression of sitting in a vehicle cockpit via haptic sensations.

The virtual cockpit was displayed in a cubic—shaped five—sided back projection system (back projection on ceiling, floor, and three walls of the cube). The length of side of the projection planes was 2.5 m each. The virtual environment ran on an SGI Onyx2 graphics engine. The participants’ head movements were tracked with a six—DOF tracker (MotionStar® by Ascension®), the left and right eye channels were separated using Stereo-Graphics® CrystalEyes® shutter glasses. The open side of the projection cube was behind participants’ backs.

All four settings were similar to those used in practical industrial applications in different contexts. As they reflect extreme examples in terms of the immersion and pictorial realism provided, they can be regarded as ‘archetype settings’.

16.2.2 Measuring presence: Which presence, and how?

From 1998 on, a number of researchers came up with multidimensional presence concepts (e.g. [1, 9, 11]). They expanded on earlier conceptualizations of the sense of presence, which were mostly focusing on the user’s sensation of (physically) being part of the virtual environment (see e.g. [12, 13], among others). In the study reported here, we were interested in the effects of immersion and pictorial realism on presence — but these effects might differ from one presence facet, i.e. dimension, to the other. A one-dimensional concept cannot reveal these possible differences.

We thus decided to employ the multidimensional concept proposed by Schubert et al. [1]. These authors suggested to differentiate between (a) ‘spatial presence’ (highly correlated to the ‘classic’ sense of being there), (b) ‘involvement’, and (c) ‘realness’. The authors claim these to be three different facets of a more general presence construct [1, 10, 14]. They base their claim on the results of factor analyses of a large data set collected in an online survey. The three presence facets correspond to factors extracted in their analyses. Their findings are generally supported by work of Lessiter et al. [11], who found similar presence facets in an extensive cross—media study.

![Figure 16.1](https://example.com/figure16.1.png)

**Figure 16.1** Views of the virtual cockpits used in the low (left) and high (right) pictorial realism settings. The cockpit on the right featured colored surface textures. The virtual steering wheel was replaced by a real one in both settings.
Measurement methods that try and deduce the sense of presence from physiological or behavioral data can hardly differentiate between the different facets of a multidimensional presence construct. Hence, we decided to employ a post-exposure questionnaire developed by Schubert et al. It is designed to measure the three facets separately. The original version of this questionnaire consists of 14 questions. Minor modifications due to the particular situation of our experiments were necessary, yielding an adapted set of 13 items (seven point Likert scale). Because of these modifications we had to perform our own factor analysis of the questionnaire results. The final questionnaire was used twice within each experimental run (versions ‘A’ and ‘B’, items identical in A and B, but in permuted order).

This enabled us to assess the reliability of our measurements.

### 16.2.3 Participants and procedure

Seventy-seven people participated in the study (nine female and sixty-eight male). Some of the participants had used virtual reality technologies beforehand, but none of them on a
regular basis. They were divided into four separate groups. Each of these groups experienced only one of our four experimental settings. All participants experienced three consecutive VE sessions. Before entering the first session, the instructor gave either a detailed description of the enabling technology (disillusioning mental priming, see Table 16.1) or an enthusiastic description of the VE as a compelling experience (mental priming enhancing the illusion).

In each session, participants were asked to perform a simple task that slightly differed between the three sessions (object comparisons). Participants in the high immersion setting were asked to perform an additional task. This task involved the interactive object scaling mentioned as an immersion factor in Table 16.1. After completion of their tasks in sessions 1 and 3, participants were asked to fill out presence questionnaires A (reading point A) and B (reading point B), respectively. While completing the questionnaires they were still sitting in the driver’s seat and were still immersed in the virtual cockpit.

16.3 Presence measurement reliability - capturing a volatile state

16.3.1 Two measurements, two (different) results

The results of questionnaires A and B were subject to separate but equally structured factor analyses (77 cases, all 13 questionnaire items used, main component analysis, Varimax rotation).

The results of questionnaire A were not suitable for factor analysis according to the measure of sampling adequacy criterion (MSA value = 0.763 < 0.8). By way of trial, we nevertheless conducted a factor analysis. It yielded four factors with eigenvalues > 1 (Kaiser-Guttman criterion), suggesting four presence facets rather than those three found by Schubert et al. [1]. In addition, the resulting factor loadings for questionnaire A did not allow a similar content-wise interpretation of the factors as in the original work by Schubert et al. Our first factor might be interpreted as their presence facet ‘realness’; but their facets ‘involvement’ and ‘spatial presence’ could not be reproduced here. Another factor analysis of this data set with a forced three-factor solution did not improve matters.

We thus excluded the results of questionnaire A from our subsequent investigation of the effects of immersion and pictorial realism on presence. In section 16.3.3, we will discuss possible reasons for the outcome of questionnaire A and compare them to those of B (remember that we used identical questionnaire items in A and B).

The data collected with questionnaire B exhibited a different structure. They proved to be suitable for factor analysis (MSA value = 0.809). The analysis yielded three factors with eigenvalues > 1. Numerically, the 13 questionnaire items could unambiguously and completely be assigned to these three factors. The assigned items allowed a straightforward interpretation of the three factors that corresponded well with that of the original authors [1].

We coined the three factors ‘reality appraisal’, ‘involvement’, and ‘spatial presence’ after Schubert et al. (they used the term ‘realness’ for the first factor [1]). Our interpretation is as follows:

- **reality appraisal** describes to what extent the participant judges the VE comparable to a corresponding real environment;
- **involvement** describes to what extent the participant’s attentional resources are directed to the VE; and
- **spatial presence** describes to what extent the participant has the sensation of being part of the VE physically.
The factor values of reality appraisal, involvement, and spatial presence for each participant were extracted by regression (from the data collected with questionnaire B only). They were normalized to a mean of zero and a variance of one. These values were used to analyze the effects of immersion and pictorial realism on presence.

16.3.2 Calculating correlations, estimating reliability

Given the striking difference of the results of questionnaires A and B, are the results of B reliable? Under certain conditions (particularly $\tau$-equivalence, see [15]), the reliability of a measurement can be assessed by calculating the correlation of corresponding data collected at different points of time (retest reliability). At first sight, the variables one should compare regarding their values at different points of time seem to be the three presence facets (extracted factors). These three — rather than the individual 13 questionnaire items — are of interest here. But, as a matter of fact, the data we collected with questionnaire A did not yield these three variables.

From our point of view, the most appropriate way to handle this problem is to use the factor coefficients generated in the analysis of the results of questionnaire B to calculate approximated factor values from the raw data (item values) of questionnaire A. By doing so, we define the factor coefficients of questionnaire B to be an integral part of our measurement tool. This seems to be somewhat justified, as only the data of questionnaire B reproduced the factor structure the questionnaire was designed for by Schubert et al. It has to be kept in mind that this is only a rough approximation, though.

Following this approach yields correlation coefficients between the data of A and B of 0.69 for reality appraisal, 0.53 for involvement, and 0.55 for spatial presence. All three correlations are statistically significant at the 0.01 level. A test of the $\tau$-equivalence of the variables suggests (but cannot prove) that these correlations can in fact be interpreted as retest reliabilities. The correlations are not particularly high, but nonetheless indicate comparable data structures in A and B. Thus, the data suggest that our presence values measured with questionnaire B are sufficiently reliable.

16.3.3 Why two results, anyway?

Why was the three-dimensional factor structure that the original questionnaire was designed for reproduced only in one of our two measurements?

Hypothesis 1. Questionnaire A was filled out only a few minutes after participants’ first VE exposure (questionnaire B at the very end of the whole session). At this point of time, they possibly had not developed a strong sense of presence yet. Consequently, the variances in the 13 questionnaire items would mainly be noise. In that case the extracted factors could not be interpreted in a sound way.

Hypothesis 2. On the other hand, it is conceivable that participants had actually developed a strong sense of presence when answering questionnaire A. But as most participants were ‘VE beginners’, the sensation of presence in a VE was an unfamiliar kind of sensation for them. Possibly they could not yet relate their impressions to the way the questionnaire requested to express them, rendering it — at that point of time — a futile tool.

In fact, both hypotheses question the interpretation of the correlation coefficients as retest reliabilities (see section 16.3.2). These coefficients are meant to compare our measurements of the users’ cognitive states in reading points A and B. According to the hypotheses, the cognitive state would either hardly have existed in reading point A (hypothesis 1), or it was not measurable with our questionnaire in A (hypothesis 2). In any case though, our results suggest that the duration of VE exposure before presence measurements
are taken has to be considered carefully. Short exposures might strongly distort presence data.

16.4 Susceptibility of presence — on the need to differentiate

16.4.1 Effects of immersion and pictorial realism on presence

The factor values — i.e., values of the presence facets — of the participants were averaged separately for each setting. Figure 16.2 displays the mean factor values and standard errors of the means. The mean values were subject to separate variance analyses (one independent variable). Due to the high noise expected in presence values, we employed a significance level of 0.1. The results of the variance analyses were the following (see also Figure 16.2):

**Reality appraisal.** The variance analyses show a significant dependence of reality appraisal on the immersion setting for high pictorial realism (PR) \( (F(1, 34) = 5.13, p = 0.030; \) higher reality appraisal for high immersion). For low PR no significant influence of the immersion setting was found. In addition, a significant dependence of reality appraisal on PR could be detected \( (F(1, 31) = 3.04, p = 0.091; \) for low immersion only). Interestingly, reality appraisal was lower in the higher PR setting. No other effects were significant.

**Involvement.** The variance analyses show a significant dependence of involvement on the immersion setting for low PR \( (F(1, 37) = 4.00, p = 0.053; \) involvement higher for high immersion). For high PR no significant influence of the immersion setting could be detected, and no other effects were significant.

**Spatial presence.** No significant effects were found.

16.4.2 Explanations

Obviously, the effects of immersion and pictorial realism on the three presence facets are not homogeneous. How can these results be understood? We propose a set of explanations that pay regard to possible interactions of the effects of immersion and pictorial realism.

(a) **Effects on participants’ reality appraisal.** As expected, lowering immersion decreased participants’ reality appraisal (from setting 4 to 2 in Figure 16.2). The immersion factor with the strongest influence on reality appraisal was probably the frame rate. In the low immersion setting with low frame rate, the virtual cockpit appeared strongly distorted when users moved their heads. The realistic impression of the high pictorial realism cockpit was partly destroyed. This probably resulted in low reality appraisal.

![Figure 16.2](image-url)  
**Figure 16.2.** Mean factor values and standard errors of the means for participants’ reality appraisal, involvement, and spatial presence in the four experimental settings (different combinations of immersion and pictorial realism). The brackets denote significant differences of the means. The numbers in black dots mark the four experimental settings and refer to Figure 16.3.
Participants’ reality appraisal values in setting 2 could be interpreted as a reaction to being ‘disappointed’: Without moving, the cockpit with high pictorial realism looked fascinating; upon head movement, participants were disappointed because the system could not maintain the illusion. Participants reacted to this mismatch with low reality appraisal assessments. This could also explain the significant increase of their assessments when lowering pictorial realism (from setting 2 to 1). In setting 1, the strong mismatch of high pictorial realism and lacking system power (i.e., low frame rate) that resulted in the disappointment was removed.

But why did the variation of the immersion have no effect on participants’ reality appraisal in the low pictorial realism cockpit (settings 3 and 1)? Probably, these cockpits were just too far from being ‘like the real thing’ (no colors, no textures applied). Here, the effect of the immersion on participants’ reality appraisal was dominated by the low pictorial realism. Their reality appraisal was ‘locked’ (see the left side of Figure 16.3).

(b) Effects on participants’ involvement. Participants’ involvement decreased when immersion was lowered (from setting 3 to 1 in Figure 16.2). Again, this might be due to the lower frame rate in setting 1. The effect was probably enhanced by the lack of interaction and the other low values of the immersion factors (see Table 16.1). But immersion affected involvement only in the low pictorial realism cockpit. Apparently, participants’ attention was strongly captured by the looks of the high pictorial realism cockpit: Their involvement was ‘locked’ for high pictorial realism (see the left side of Figure 16.3). In other words, the influence of pictorial realism dominated that of immersion.

(c) Effects on participants’ spatial presence. No influence on participants’ spatial presence was detected. The systematic effects of immersion and pictorial realism on all three presence facets were overlaid by participants’ individual reactions to the offered stimuli.

Strong influence of individual user characteristics on the sense of presence has been proposed by various researchers (see e.g. [7, 10]). One of these characteristics is the users’ willingness to suspend disbelief [8]: The willingness to overlook the fact that the experience is mediated by technology. This willingness is essential for the development of spatial presence — but it is merely helpful for increasing reality appraisal and involvement. Thus, random variations of this willingness among participants have overlaid systematic effects of immersion and pictorial realism most strongly on spatial presence.

Figure 16.3. Left: Partial dominance of pictorial realism (PR) over immersion regarding their effects on two presence facets (‘locked’ at mean values, see text). The numbers in black dots refer to the four experimental settings. Right: The four settings are related to different costs involved in their generation.

<table>
<thead>
<tr>
<th></th>
<th>immersion &amp; system costs</th>
<th>pictorial realism &amp; labor costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>2</td>
<td>low</td>
<td>high</td>
</tr>
<tr>
<td>3</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>4</td>
<td>high</td>
<td>high</td>
</tr>
</tbody>
</table>
16.4.3 Designers and treasurers, watch out!

These results should be taken into consideration when designing a VE that is to generate a certain degree of presence. Setting up a VE with high immersion — particularly realizing high frame rates regardless of scene content — usually involves high system costs. On the other hand, achieving high pictorial realism often involves high labor costs (construction of detailed 3D geometry, generation of realistic textures etc.).

Thus, our results might be regarded as a further step towards mapping certain cost types (and preparation time) involved in generating a VE to the achievable effects on the users’ sense of presence (see the right side of Figure 16.3). If e.g. for a certain application high reality appraisal is crucial, pictorial realism should apparently be high. But the high labor costs involved in its generation are not well spent if immersion is too low: According to our findings, a mismatch of pictorial realism and system power might lead to a surprisingly low reality appraisal. Low system power, combined with low pictorial realism, might prove to be a better compromise then.

16.5 Some implications for the conceptualization of presence

To contemplate the nature of the sense of presence is a difficult, often philosophical undertaking. Nonetheless, our empirical results might add some aspects to the discussion.

First of all: You tend to get what you ask for. There is always a certain degree of subjectivity in developing a questionnaire. The one used here is designed to measure a three-dimensional presence construct. Nevertheless, the different effects of the setting variations on the three presence facets recorded in this study strongly suggest the appropriateness of a multidimensional presence concept. As a matter of fact, this notion is supported by empirical results exhibiting heterogeneous effects of these three facets on another cognitive parameter, notably spatial perception. Hofmann et al. showed that rising involvement and spatial presence had effects on spatial perception that were directly opposed to those of rising reality appraisal [16].

Having said that, these very results could be interpreted in a diametrical fashion: One could argue that only one of the three facets measured here is ‘the proper sense of presence’, the others being unconnected cognitive phenomena. But then again, it is hard to argue against the assertion that the degree of selective attention to the VE (i.e., involvement) was an integral part of the ‘sense of being there’.

Finally, there might well be facets of the seemingly complex cognitive state of presence we have not even started to consider. A lot of research remains to be done in the field of presence — further empirical work is particularly needed.

16.6 Conclusions

In the study reported in this chapter, we measured the effects of immersion and pictorial realism on three facets of the sense of presence in a virtual environment. Presence was measured twice, using the same set of questionnaire items. The presence data proved to be sufficiently reliable. However, our results demonstrated that measuring presence is a delicate and time-critical business. With regard to the underlying conceptualization of the sense of presence, our experiments support the applicability and usefulness of a multidimensional presence construct.

Concerning the susceptibility of presence (or more precisely, that of our definition of presence), we detected effects of the variation of immersion on two of the three presence
facets analyzed here: reality appraisal and involvement. Both were raised by an increase of immersion, as common sense would suggest. But this effect was strongly and selectively dominated by the influence of pictorial realism. Pictorial realism ‘locked’ reality appraisal and involvement at mean values, depending on its degree. In other words, the influence of immersion was disabled. The third presence facet analyzed here — spatial presence — was partly not affected by setting variations. We proposed explanations for the effects observed.

Based on these explanations, we indicated practical implications of our results for the design of VEs. Provided that our findings were combined with a bulk of similar research work yet to be done, a mapping of the resources needed for setting up a VE and the achievable effects on the user’s sense of presence could be established. This would facilitate the effective allocation of efforts and funds involved in designing virtual environments for practical tasks — if, and only if, further research will be done to identify the application areas in which a strong sense of presence is beneficial at all.

16.7 References


