

PRESENCE 2008

**Proceedings of the 11th Annual International Workshop on Presence
Padova, 16-18 October 2008**

Printed by
CLEUP Cooperativa Libreria Universitaria Padova
Padova 2008

Edited by Anna Spagnolli, Luciano Gamberini

ISBN: 978-88-6129-287-1

© The copyright for this publication as a whole stands with HTLab. The copyright of each separate paper published within these proceedings remains vested in its author. Authors have assigned to Presence 2008 organizers and ISPR (International Society for Presence Research) the on demand availability rights for their work and the right to create a derivative work from it, including publication on ISPR website and on the conference proceedings.

When and How to Assess Subjective Overall Judgments of Presence?

Bartholomäus Wissmath, David Weibel, Daniel Stricker

University of Bern, Switzerland

Swiss University Institute of Distance Education

{ wissmath@psy.unibe.ch, david.weibel@psy.unibe.ch, daniel.stricker@psy.unibe.ch }

Abstract

This study investigates differences between subjective on-line- and post-immersion measures, verbally and pictorially anchored scales, and the effects of content on those different measures. These factors were investigated by means of a 2x2x2 within-subjects-design. Participants (N = 162) evaluated two video clips. Against our expectations the findings suggest on-line- and post-immersion measures to be interchangeable. In line with findings from other fields than presence, pictorially anchored items seem to have major advantages when overall judgments are to be assessed. The advantages of pictorially anchored items apply in particular for language-containing environments.

Keywords--- On-line Measurement, Post-Rating, Verbal Measures, Pictorial Measures, SAM, Telepresence

1. Introduction

Since Marvin Minsky [1] coined the term Telepresence, an abundance of conceptualizations have been proposed by scholars of various disciplines. According to Lombard and Ditton [2] these conceptualizations include social richness, realism, transportation, immersion, social actor within medium, and medium within a social actor. These conceptualizations can be further grouped into the two main categories physical and social presence. Thus, there is still no general conceptualization of presence and the community is still challenged to refine and standardize presence definitions [3].

2. The measurement of presence

Correspondingly, a commonly accepted paradigm for the assessment of presence does not exist yet. Though the ideal instrument to assess presence is not known so far, the desired features of such an instrument have been described a decade ago [4]: Relevance, validity, reliability, sensitivity, non-intrusiveness, robustness, and convenience.

In their compendium, van Baren and IJsselsteijn [5] provide an overview of presence measures. Those instruments can be grouped into objective and subjective measures. The objective assessment of presence consists of either behavioral

or physiological measures. Thereby the major drawback is that the collection of these data is often intrusive and sometimes unreliable [5]. In addition, the relation between these measures and the level of presence is in most cases not necessarily strong [5].

2.1. Subjective measures

In this study, we focus on subjective measures. Sheridan [6] argues that the subjective experience of presence is ideally measured by means of subjective assessment. These indicators can be further categorized into qualitative measures, presence questionnaires, continuous assessment, psychophysical measures, and subjective corroborative measures [5]. Presence questionnaires are usually administered after the exposure and are widely used. An indication for the dominance of post rating measures can be found in the presence measurement compendium [5], which lists 29 subjective measures [e.g. 7, 8]. Of those, only one single instrument is intended for continuous assessment while the 28 other measures are post-rating instruments. According to Insko [9] key advantages of these post-rating questionnaires are ease to administer, high face validity, and the lack of measurement related interferences during exposure. Additional benefits of questionnaires are the opportunity to conduct factor or cluster analysis, which allow the identification of the underlying dimensions of presence, low cost, mobility, sensitivity, and ease to analyze and interpret [5]. These advantages made subjective post-rating scales the most used presence measure.

There is no powerful measure without side effects or drawbacks. For Sadowski and Stanney [10] a major obstacle is to assure validity as participants must understand the concept of what presence is and interpret questions uniformly. According to Insko [9], further disadvantages associated with post-immersion questionnaires are anchoring effects, inaccurate recall, and inability to assess temporal variations in the subjective sense of presence. Van Baren and IJsselsteijn [5] underline possible recency effects in the post-rating judgments.

2.2. Continuous assessment

To overcome these limitations, continuous assessment has been proposed [11, 12]. This allows the assessment of

variations in the subjective experience of presence, which are likely to occur not only through changes in the stimulus but also through the participant (e.g. increasing fatigue during exposure) and to overcome limitations associated with post-rating measures [12]. A method originally developed to assess picture quality of TV images was adopted to continuously assess the experience of presence during exposure. Therefore participants had to provide on-line judgments of presence by means of a hand-held slider. A task the authors consider to require little attention and effort to operate. For non-interactive stimuli a considerable temporal variation depending on the sensory input was found [11].

A drawback of this procedure may be that participants are required to divide their attention between the physical and the mediated environment. Attention allocation towards the display is a central component of spatial presence [13]. Correspondingly, Riva, Waterworth, and Waterworth have suggested that is the result of the unification of proto, core, and extended awareness in the mediated environment [14]. Thus, on-line ratings could restrain participants to experience the arrival in the mediated environment as well as departure from the actual physical setting. Both of these factors were found to be central components of presence [15]. In addition, one could expect that a person who is fully present in the displayed environment not only forgets about the real world but also forgets to shift the slider to “fully immersed”. Not only the reliability but also the validity of this method is to question as during exposure participants could confuse presence with other perhaps more familiar judgments such as liking or enjoyment.

Ijsselstein, De Ridder, Freeman, and Avons [12] consider on-line assessment of presence to be mainly applicable to non-interactive media as continuous measurement devices may interfere with operating an interactive environment.

The advantages and drawbacks of both, on-line judgments as well as post-ratings have been described above. To our knowledge, a direct and systematic comparison of on-line- and post-exposure methods to assess the subjective sense of presence does not exist so far. Based on the theory and the findings reported above, the following hypothesis is offered:

H1: The on-line assessment of presence draws the allocation of attention away from the medium and thus leads to lower levels of presence.

2.3. Verbal vs. non-verbal subjective assessment

Various non-verbal methods to assess presence have been suggested in the past [9, 12]. Thereby the main approaches are psychophysical or subjective corroborative indicators. The former include cross-modality-matching (CMM), free-modulus magnitude estimation, paired comparison, and the virtual reality Turing test. Although such measures were found to be sensitive, cheap, unobtrusively, and easy to use, only a limited number of presence studies have been adopting such measures [5, 12]. This could be due to the fact that these measures are

prone to bias as they rely on the experimenter’s instruction and the subjective interpretation of what to rate.

Subjective corroborative measures do not assess presence directly but aim on psychological processes, which are conceptually related to presence. Among those, time estimation, attention allocation, recognition, recollection, and spatial cognition [5] have been adopted. These indicators may serve as unobtrusive measures but they assess at best concepts, which are closely related to presence but not presence directly.

However, the majority of subjective presence instruments are based on verbal measures. The advantages of the verbally based subjective measures such as high face validity or applicability to almost each mediated environment might explain the dominance of the verbal indicators. In most cases, participants rate statements or questions assessing the experience of presence on Likert-type scales.

In contrast to those verbal based measures, visually oriented scales are supposed to have the potential to be culture free [16]. They are easy to use and to understand—even for children and people who cannot read. Another advantage in comparison to verbal measures is that participants respond them quicker and usually enjoy those items more [16]. In addition, depending on the sensory modalities of the mediated environments and the resulting differences in the mental workload [17], verbal and non-verbal measures could be unequally appropriate. When measuring presence on-line in visually dominated environments, verbal measures could be more accurate. For environments in which language based information is of major importance, verbal measures could interfere information processing more than a visual measure and thus prevent participants from the illusion of non-mediation. In the development of the Temple Media Questionnaire [18] a Self Assessment Manikin (SAM) measure [19] was introduced to the presence measurement. The pictorially anchored presence assessment technique starts with the following instruction [18]:

Please use the figures below to indicate your feelings or emotional response to the media experience. The pictures go from a person who feels he or she is **INSIDE THE PICTURE, A PART OF THE STORY, A PART OF THE ACTION** on the left end, to a person who feels he or she is **OUTSIDE THE PICTURE, REMOVED OR SEPARATED FROM THE STORY, NOT PART OF THE ACTION** on the right end. Please put an ‘X’ through the picture, or in the space between any two pictures, that best represents how you felt during the media experience.

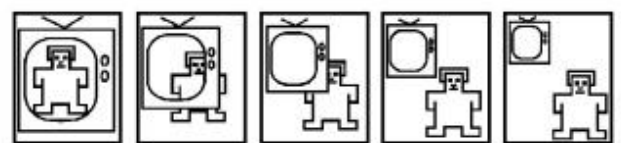


Figure 1 Depiction of the pictorial item

Based on the findings described above, the following hypotheses are proposed:

H2: Compared to verbal based measures, visually oriented scales will allow higher levels of presence.

H3: When adopted on-line, verbal measures of presence interfere the sensation of presence in language-based environments, whereas visual oriented scales interfere the illusion of non-mediation in visually dominated environments.

3. Determinants of presence

The examination of presence measures is especially fruitful when we have the central factors in mind, which potentially influence the sense of presence. Ijsselstein, De Ridder, Freeman, and Avons [12] reviewed various theoretical analyses and concluded that four factors contribute to the sensation of presence: First the *extent and fidelity of sensory information* which include resolution or spatialized audio. The second factor refers to the *matching between sensors and display*. The third category *content factor* is very broad and includes the objects, actors and events presented by mediated environment. The fourth factor consists of *user characteristics* such as cognitive resources, interest or experience with a particular medium. From our point of view, all factors besides the second apply to interactive as well as to non-interactive media. Television and video clips are clearly to consider as non-interactive, but they elicit sensations of presence [15, 20, 21]. In the light of those findings, the following hypothesis is suggested:

H4: The content of the medium will influence the levels of presence.

4. Method

The overall design was a 2x2x2 factorial within-subjects-design. Factors were time of measurement (levels: on-line- vs. post-exposure), type of measure (levels: language free vs. language based), and audio content (levels: instrumental music vs. language). Participants (N = 162) watched two video clips.

4.1. Stimuli

The two film clips used for this experiment were two 4-minutes lasting sequences of television broadcasts. The language-free clip showed the planet earth. The language based clip dealt with the perils of global warming. An environmental campaigner was interviewed by a journalist before and while they made a trip in an off-road vehicle.



Figure 2 Screenshots of the stimuli

4.2. Independent variables and measures

The first factor was *time of measurement*. It was manipulated by having participants in one condition to give three on-line ratings on single item measures every 60 seconds, starting from 30 sec. after the clip had started. In the other condition, participants evaluated their sense of presence on a single item post-exposure measure.

The second factor was *type of measure*. In one condition participants evaluated their feelings of presence on the following verbally coded single item: “To what extent do you feel located in the world displayed by the clip?” (5-point rating from “very much” to “not at all”). In the other condition the pictorial item [18, 19] described above was administered (cp. Figure 1).

The third factor was *content*. In the non-language condition film sequences from the space shuttle (cp. Figure 2) dubbed with instrumental music were presented. In the language condition the interviewer and the environmental campaigner discussed the perils of global warming, the future of the planet, and human behavior when facing global dangers.

4.3. Participants

For this experiment undergraduate students enrolled in psychology were recruited. One hundred and sixty-two undergraduate students volunteered to participate in this investigation. Among those the majority was female (82.7%).

4.4. Procedure

Participants were assigned in groups to one of the eight experimental conditions. The experiments were conducted in standard lecture rooms. The alignment of the chairs was similar to the setting in a movie theater. The stimuli were presented by means of high definition video beamers. Equal projection size and audio levels for all conditions were ensured. Each participant saw both clips. The questionnaires were provided in paper-pencil form and were handed over before the experiment began. The experimenter told the participants that there were no wrong answers and that they could end participation at any time. Before starting the clip to be on-line rated, the experimenter asked the participants to examine the corresponding instructions and items and to answer one item each time he would clap his hands. The post-rating measures

were mentioned only after the presentation of the clip to be rated.

During the experiment, each participant rated the sense of presence for one clip by means of the verbal and the other clip by means of the pictorial measure. Each participant rated one clip on-line (3 judgments, 60 seconds time lag between measures, the answer was triggered by a handclapping of the experimenter) and one clip post-exposure. To prevent sequence-effects, the order of the within-factor levels was counterbalanced. The whole experiment took about 10 minutes.

5. Results

All hypothesized have been analyzed by means of a three-factorial linear mixed model. Hypothesis 1 predicting the on-line assessment ($M = 2.03$; $SD = 1.40$) of presence leading to lower levels of presence than the post rating ($M = 2.05$; $SD = 1.56$) could not be supported at all, $F(1, 158) = 0.01$, $p = .91$.

In contrast, the second hypothesis was strongly supported. As predicted, visually oriented scales ($M = 2.18$; $SD = 1.48$) result in higher levels of presence compared to verbally based measures ($M = 1.90$; $SD = 1.47$), $F(1, 158) = 9.60$, $p = .00$.

The third hypothesis was partially supported. There is an interaction between type of measure and content, $F(1, 158) = 7.74$, $p = .01$. Figure 4 shows that text based measures interfere presence in language-dominated environments. In contrast to that, visual measures seem to be equally suited for both types contents.

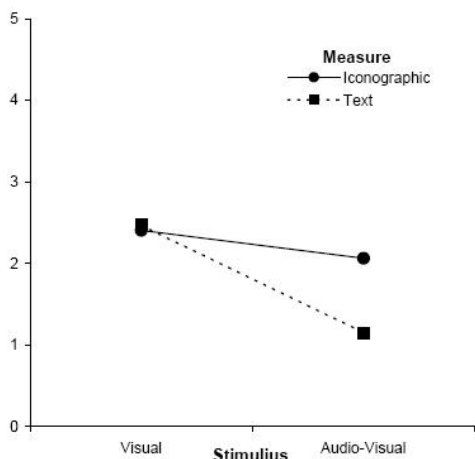


Figure 3 Interaction between content and measure

Hypothesis 4 was also supported. The content of the medium influences the levels of presence as the language-free condition ($M = 1.64$; $SD = 1.38$) produced lower presence than the language-based clip ($M = 2.44$; $SD = 1.48$), $F(1, 158) = 36.92$, $p < .000$.

For exploratory reasons, the interactions time of measurement x content, and time of measurement x type of measurement x content were tested. Both were non-significant

with $F(1, 158) = 3.03$, $p > .08$, and $F(1, 158) = 3.50$, $p > .06$, respectively.

6. Discussion

The question whether on-line and post-rating measures are interchangeable has been raised various times [5, 11, 12]. So far, scholars who investigated experiences of presence by means of subjective post-rating measures could not be sure that these widely used post-experience judgments are reliable and valid indicators of the presence-process. Therefore one of the most noteworthy findings of this investigation is that averaged on-line judgments of presence and post-rating measures lead to identical results. At least for rather short exposure times, participants seem to provide highly accurate judgments of their over-all presence experience. This finding could be a strong point for the dominance of post-rating scales. On the other hand, our findings suggest not to worrying that subjective on-line ratings draw a meaningful portion of the attention away from the displayed environment and thus diminish the sensation of presence. However, these results cannot be directly applied to paradigms consisting of more than single-item measures, longer or shorter exposure duration and to more interactive or complex environments.

Another noteworthy finding is that visually oriented scales lead to higher ratings of presence. This could be because they are easier to understand and quicker to use than verbally anchored items. Another explanation could be that participants enjoy those items more [16] and therefore give higher presence ratings. However, a problem associated with visually oriented items could be increased item-complexity. In our case, a single overall item proved as valid and reliable indicator of presence. To what extent this may be true for more complex sub-dimensions of presence such as interface awareness or suspension of disbelief is not clear so far.

We expected, depending on the sensory modalities of the mediated environments, verbal and non-verbal measures to be unequally appropriate. In line with our prediction, text based measures actually seem to interfere the sensation of presence. This could be due to impaired information processing as the mental workload during responding verbally anchored items is higher [16]. This seems to prevent participants from the illusion of non-mediation. However, for visually dominated environments this seems not to be true as both measures provide identical ratings. A possible explanation in favor of visually anchored measures is that the mental workload to respond them is so small that the allocation of attention towards the mediated environment is not impaired regardless of the content.

Not surprisingly, the content of the displayed clip influenced the presence ratings. As the clips differed in many aspects associated with the content such as topic, display of humans, editing, location and the like, there is no way to relate the differences in the presence assessment specific features of the clips. Still we consider the effect of this factor to be meaningful not only because it adds evidence to the growing

body of research demonstrating that content factors matter [12], but because it increases our confidence to believe in the interchangeability of on-line- and post-rating measures. In combination with the variance obtained in all measures, this effect provides a strong point that our participants experienced different levels of presence in the course of the experiment. Therefore the identical results for on-line- and post-rating measures cannot be explained with low variance or response biases, which in turn strengthens our point.

To ensure the validity of the findings discussed above, we made sure that our findings concerning time of measurement equally apply for different contents. The corresponding interaction clearly failed to reach significance but the alpha size is less convincing than in the case of hypothesis 1. To appraise this finding it is important to mention that our design includes a large sample, repeated measurement, and linear mixed modeling. These three features contribute to high power, which in turn provides strong support for all of our findings, especially when we refer to the interchangeability of measures suggested in the discussion of the first hypothesis.

7. Limitations

The present study has a number of limitations and their identification should help to refine future research efforts. For this study, a student population was tested under experimental conditions. Clearly, a more representative sample as well as the replication of these findings for other measures would increase confidence in the results. In addition, our findings are based on non-interactive environments and overall presence-judgments. Having adopted single-items measures, the impact of time and type of measurement on the sup-dimensions of presence has not been addressed here. Before directly generalizing our findings to more interactive environments, longer or shorter exposure times, and other presence measures, additional research is required.

Conclusions

This piece of research suggests the adoption of post-rating scales as participants seem to be able to provide ex post highly accurate overall estimation of the presence experienced. However, when temporal variations of presence are of particular interest, our study suggests that overall on-line ratings do not interfere the sense of presence.

This study contributes to findings suggesting advantages of visually anchored measures in terms of efficacy and validity. The presence community could try to establish and investigate more specific (i.e. addressing sub-dimensions of presence) non-verbal subjective rating tools.

When assessing an overall estimation of presence, the advantages of visually anchored measures seem to pay off especially with language-based environments. Researchers should keep that in mind when setting up an investigation focusing on such environments.

References

- [1] M. Minsky. Telepresence. *Omni*, 2, 45-51. 1980.
- [2] M. Lombard, T. B. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-mediated communication*, 3, 1997. URL: <http://209.130.1.169/jcmc/vol3/issue2/lombard.html>.
- [3] M. Lombard. Presence and telepresence scholarship: Challenges ahead. Keynote presentation at *When Media Environments Become Real (WMEBR)*, University of Bern, Switzerland. February. 2008.
- [4] C. Hendrix, W. Barfield. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, 5, 274-289. 1996.
- [5] J. van Baren, W. A. IJsselsteijn. *Measuring presence: A guide to current measurement approaches*. Deliverable 5, IST-FET OMNIPRES Project. 2004. URL: <http://www.presence-research.org/>
- [6] T. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1, 120-126. 1992.
- [7] B. Witmer, M. Singer. Measuring presence in virtual environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7, 225-240. 1998.
- [8] M. Slater. Measuring Presence: A Response to the Witmer and Singer Questionnaire. *Presence: Teleoperators and Virtual Environments*, 8, 560-566. 1999.
- [9] B. Insko. Measuring presence: Subjective, behavioral and physiological methods. In: G. Riva, F. Davide, W. A. IJsselsteijn (Eds.) *Being there: Concepts, effects, and measurement of user presence in synthetic environments*. The Netherlands: IOS Press, Amsterdam. pp. 109-119. 2003.
- [10] W. Sadowsky, K. Stanney. Measuring and managing presence in virtual environments. In: K.M. Stanney (Ed.) *Handbook of Virtual Environments Technology*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. pp. 791-806. 2002.
- [11] W. A. IJsselsteijn, H. De Ridder, R. Hamberg, D. Bouwhuis, J. Freeman. Perceived depth and the feeling of presence in 3DTV. *Displays*, 18, 207-214. 1998.
- [12] W. A. IJsselsteijn, H. De Ridder, J. Freeman, S. E. Avons. Presence: Concept, determinants and measurement. *Proceedings of the SPIE*, 3959, 520-529. January. 2000.
- [13] W. Wirth, T. Hartmann, S. Böcking, P. Vorderer, C. Klimmt, H. Schramm, T. Saari, J. Laarni, N. Ravaja, F. Ribeiro Gouveia, F. Biocca, A. Sacau, L. Jäncke, T. Baumgartner, P. Jäncke. A process model of the formation of Spatial Presence experiences. *Media Psychology*, 9, 493-525. 2007.
- [14] G. Riva, J. A. Waterworth, E. L. Waterworth. The Layers of Presence: a bio-cultural approach to understanding presence in natural and mediated environments. *Cyberpsychology and Behavior*, 7, 402-416. 2004.
- [15] T. Kim, F. Biocca. Telepresence via television: Two dimensions of telepresence may have different connections on memory and persuasion. *Journal of Computer Mediated Communication*, 3, 1997. URL: <http://jcmc.indiana.edu/vol3/issue2/kim.html>.
- [16] P. Lang. *The cognitive psychophysiology of emotion: anxiety and the anxiety disorders*. Hillsdale NJ: Lawrence Erlbaum. 1985.
- [17] H. Jex. Measuring mental workload: Problems, progress, and promises. In: P. A. Hancock, N. Meshkati (Eds.) *Human Mental*

- Workload*. Amsterdam, Netherlands: North-Holland. pp. 5-29. 1988.
- [18] M. Lombard. Temple Presence Inventory (TPI) questionnaire items and factor structures. 2005. URL: <http://astro.temple.edu/~lombard/research/P2qstnre.doc>.
- [19] A. Lang. Self Assessment Manikin (SAM) measure. 2000. URL: http://www.temple.edu/ispr/frame_measure_t.htm.
- [20] C. C. Bracken. Presence and image quality: The case of high definition television. *Media Psychology*, 7, 191-205. 2005.
- [21] M. Lombard, R. D. Reich, M. E. Grabe, C. C. Bracken, T. B. Ditton. Presence and television: The role of screen size. *Human Communication Research*, 26, 75-98. 2000.