

The Visual, the Auditory and the Haptic – A User Study on Combining Modalities in Virtual Worlds

Julia Fröhlich and Ipke Wachsmuth

AI & VR Lab, Faculty of Technology, Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
{jfroehli, ipke}@techfak.uni-bielefeld.de

Abstract. In order to make a step further towards understanding the impact of multi-modal stimuli in Virtual Reality we conducted a user study with 80 participants performing tasks in a virtual pit environment. Participants were divided into four groups, each presented a different combination of multi-sensory stimuli. Those included real-time 3D graphics, audio stimuli (ambient, static and event sounds), and haptics consisting of wind and tactile feedback when touching objects. A presence questionnaire was used to evaluate subjectively reported presence on the one hand, and on the other physiological sensors were used to measure heart rate and skin conductance as an objective measure. Results strongly indicate that an increase of modalities does not automatically result in an increase of presence.

Keywords: Presence, User Study, Multi-modal Feedback, Virtual Reality

1 Introduction

Ever since Morton Heilig developed the Sensorama Simulator [8], multi-sensory feedback has been claimed to be of notable importance. Half a century later the quality of graphical presentation has increased dramatically, but many virtual reality applications fall short on presenting multi-sensory experiences. Even worse, the stimuli present are sometimes in conflict with the virtual world (e.g. a silent virtual room but the air condition in the Lab is noisy).

One of the major goals in Virtual Reality is to create a highly immersive environment. Modern hardware facilitates real-time 3D graphics. In 3D-setups like the CAVE [3], the user is located directly in the virtual world and becomes part of it. However, there are other factors which influence the user's immersion, for instance natural interaction and navigation. Moreover the feeling of being in the world is also influenced by the way the user experiences the world with other senses. In order to make a further step towards understanding the correlation between multi-sensory stimuli and the perceived presence we conducted a user study.

In the following we start by giving an overview of related work. Our focus will be on research of immersion and presence tied to Virtual Reality applications and multi-sensory stimuli and their impact. The second part of the paper will describe the conducted user study, starting with an overview of the setup and procedure, followed by results. A discussion and some directions for future work will sum up our contribution.

2 Related Work

Presence has been defined as “a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences” [18]. An immersive virtual world helps users to accomplish their tasks in an efficient way: it facilitates building a mental model of the environment [13]. Moreover, existing mental models of interaction in the real world can quickly be adapted to those needed in the virtual environment. There are different opinions about how to maximize immersion. Sheridan suggested three essential factors as follows [15]:

1. The quality (and quantity) of visual, auditive and haptic feedback
2. The possibility of moving the point-of-view and the sensors in a virtual environment
3. The possibility of changing the environment, as easy as in the real world

The second factor can be regarded as accomplished. By combining a CAVE [3] and a tracking system, the user is able to move around freely (in the limited space of the CAVE) and the viewpoint adapts to his position in real-time. However, the first and third factor are only partly accomplished. Visual quality of immersive worlds is at a point of nearly being photo-realistic. Other modalities like auditive and haptic feedback are continuously enhanced, but still, in many cases there is a lack of multi-modality.

Acoustic and tactile feedback are used more commonly and there are many different concepts and devices. The integration of wind is not as common. An additional wind setup was e.g. implemented by Deligiannidis and Jacob [4]. It was used to improve speed perception as the user was navigating in a 3D-world with a scooter. Since they had a fixed wind direction their setup was limited to specific scenarios. But yet, they conducted a user study which proved not only a higher reported presence but an objectively better task performance.

Measuring presence is still a challenging task and many different types of experimental setups have been proposed. The most common measure is reported presence through questionnaires. Over the years several questionnaires have been developed, in particular the Witmer-Singer [18] and Slater-Usuh-Steed [17] questionnaires have been used in numerous studies. Those questionnaires are considered *subjective* measures, because different persons may respond totally different to the same environment. Therefore researchers are looking for more *objective* measures as well. Physiological reactions (heart rate, skin conductance, and skin temperature) were tested in virtual environments in order to find a correlation

with reported presence [10]. For certain stress-inducing environments this correlation was significant. From the results of the visual cliff studies by Gibson and Walk [7] in 1960 the idea of a virtual pit evolved. The initial experiments showed that the presence of a cliff is a fear-evoking experience and for most people it requires a huge amount of willpower to cross a precipice like this. First introduced by Mel Slater and colleagues in 1995 [16] – virtual pits are a commonly used test scenario in virtual worlds today. They evoke a physiological reaction and therefore facilitate the availability of an objective measure of presence.

To investigate which influence further modalities have on the perceived presence, Dinh et al. [5] conducted a user study in 1999. It indicated that an increased amount of modalities results in increased perceived presence and memory of objects in the environment. The environment was presented with different combinations of multi-sensory stimuli consisting of head-mounted graphic display, auditory, tactile, and olfactory cues. Each modality had two levels of realism. Results showed significant effects for auditory and for tactile cues. For olfactory cues a non significant trend was measured. Surprisingly the quality of the visual cues had no impact on the perceived presence. The authors argued that additional sensory cues, except for visual ones, work in a simple additive fashion on the sense of presence. Whether this still holds for today's virtual reality applications, is a subject of the study presented in the sections to follow.

3 Experimental Setup

In this section we describe the experimental setup used for the study. First we present the hard- and software setup followed by a discussion of the used navigation and interaction method. Furthermore a description of the virtual world as presented in the study is given.

3.1 Setup

Our setup consists of a 3-sided CAVE-like environment. The user wears tracked glasses for dynamical adaptation of the viewpoint. Furthermore a sound and a wind setup are employed to generate multi-sensory stimuli.

Spatial sound is realized with eight speakers (one in each corner) and two subwoofers underneath the floor. Sounds are divided into three different types. *Ambient sounds* represent a base level of output which is more or less constant over a larger region of a scene. As long as the user is within the defined area, ambient sound will be played without direction and always at the same volume. In addition, this concept allows the definition of environmental properties which influence the audio rendering, to fit the environment, such as an outside scenario, a cave or a concert hall. *Static sounds* are directly coupled to an object. They are adjusted in volume and direction with regard to their position relative to the user. *Event sounds* are only triggered when the related event occurs (e.g. a ball hitting the floor).

Tactile feedback is accomplished by ART fingertracking devices which track the thumb, the index finger and the middle finger. At the tip of each finger three wires made of memory metal are attached which shorten when heated momentarily. When repeated in short intervals a vibration is created, which can be utilized as haptic feedback to the user. The strength of feedback can be regulated steplessly.

Wind effects are accomplished by eight controllable fans, which are located at the upper bound of the projection area. In consideration of available space, costs, as well as fine-grained adaptation of wind direction we chose a setup in which the fans are mounted evenly distributed on a nearly circular arrangement. The fans were chosen with a special focus on being as silent as possible.

3.2 Navigation and Interaction

There are many different interaction methods in Virtual Reality, but most of them aim for efficiency instead of realism. The possibility of fast manipulation of objects is often more important than intuitiveness and ease of use. Still, a user will naturally try to grasp an object directly. After manipulating, the user will expect the object to fall down on the ground, like in the real world. That is why a natural hand interaction method was chosen in our scenario.

In most CAVE setups realistic walking is not feasible. Thus, it has to be replaced by a less intuitive navigation metaphor, but still aiming for increased immersion. A study by Slater et al. [16] indicated that walking-in-place resulted in a higher subjective sense of presence than a push-button-fly (along the floor plane) navigation. Instead of tracking the users head movements to indicate whether they are walking or not, we thus track the feet directly with markers. For rotation, we make use of users' head orientation. Sudden viewpoint changes are interpreted as changes of walking direction. Similarly to the original paper [16], the viewpoint is then dynamically adjusted until the user faces the front wall again [14].

3.3 Environment

The experiment subsequently described took place in a virtual pit environment. The presented world as seen in Figure 1 was employed in the study. It consisted of two rooms, a training room (right) and the virtual pit room (left). The virtual world was designed based on *Instantreality*, a consistent, platform-independent framework for fast and efficient application development for Virtual and Augmented Reality [6].

Our test scenario is designed similar to other virtual pit environments (e.g. [11]) with some adjustments. Since we used a CAVE and not a head-mounted display (HMD) as most similar studies did, the training room was constructed bigger. This is to give users a chance to try the walking-in-place navigation. The training room was furnished as a living room and offered enough details to spend time on exploring the environment. Furthermore we supplied some objects (a ball and a gong) to train natural hand interaction. The pit room



Fig. 1. The virtual environment as presented in the user study.

was not furnished but contained only the pit with a small gallery and two planks. The pit was actually covered with a virtual glass floor, therefore it was possible to walk right over it (actually nearly nobody did this).

The presentable modalities were:

- step sounds when walking (steps on a wooden floor)
- a radio playing the theme from 'the Good, the Bad and the Ugly'
- a pretty loud event sound when hitting a gong
- drop sounds when a ball fell on the ground
- a mechanical sound when the door to the pit room was opened
- atmospheric wind sound when entering the pit room
- tactile feedback when touching objects
- haptic wind blowing from the open windows into the pit room

4 User Study

To measure the effect of the presented modalities on users' presence we conducted a study with 80 participants. Participants were recruited through postings in the university building and they were rewarded with chocolate. The only mandatory requirements were that participants had no significant fear of heights, did not participate in any of our previous studies, and were native speakers of German.

Participants were divided into four groups each presented a varied combination of modalities. As a baseline all groups were presented the same graphical world: a virtual pit setup with a training room and a pit room as described in sect. 4.3. The first group did not get any further modalities besides the graphical one, while the second group had additional acoustic feedback whereas the third group had additional haptic feedback. The fourth group was presented the full combination of visual, auditory, and haptic stimuli.

4.1 Questionnaires

There were six types of questionnaires used in this study as described in the following:

1. A questionnaire asking for demographic information.
2. The Immersion Tendency Questionnaire (ITQ) by Witmer and Singer [18]. It consists of 12 questions to measure the capability of individuals to get immersed in daily activities like reading or watching a movie.
3. The Simulator Sickness Questionnaire [9] – given before and after the study – to measure influence of the virtual trip on participants’ health condition.
4. The two height anxiety questionnaires as introduced by Cohen [2] consisting of 20 situations which evoke a fear of heights.
5. A presence questionnaire similar to the University College London (UCL) Presence Questionnaire, also known as the Slater-Usuh-Steed (SUS) Questionnaire [17]. It consists of 13 questions concerning the overall experiment.
6. A questionnaire with open questions about participants’ experience. The questions were designed to check if the presented modalities were noticed and if they were appropriate. Moreover participants had the chance to write a few lines about how they liked the experiment and if there were any improvements suggested.

4.2 Procedure

Before entering the training room participants had to answer five questionnaires: demographic information, Immersion Tendency Questionnaire, Simulator Sickness Questionnaire and the two height anxiety questionnaires. Then a calibration of markers for gesture recognition was performed to ensure the same conditions for each participant.

All participants started with a training procedure of about 15 minutes. Step-by-step, they learned to look around, to walk and to interact with objects in the virtual world. The training included a walk through the whole training room, hitting the gong and throwing a ball. Afterwards participants were asked to proceed to the pit room and throw two balls at a target on the ground of the pit. After the part within the virtual world was completed, another set of questionnaires was given to the participants: the Simulator Sickness Questionnaire as before, the UCL Presence Questionnaire, and the questionnaire with open questions about their experiences including memory questions. In addition we recorded physiological data through heart rate and skin conductance sensors, in order to measure the physical reaction to the virtual pit objectively. The whole procedure took about 60 minutes for each participant.

5 Results

As described above there were two measures of presence: subjective and objective. First we will present the results of the subjective measure from the UCL

questionnaire, which consists of 13 questions. Participants were to answer the questions on a 7-point Likert scale. The topics covered are: the sense of being in the Virtual Environment (VE), the extent to which the VE becomes the dominant reality, and the extent to which the VE is remembered as a place. Three of those 13 questions are used to measure the reported *behavioral presence* based on studies indicating that behaviors as a response to stressful stimuli and reported behavioral presence correlate. Therefore a participant would react with more pit-avoidance, the more real the precipice would seem. Another three questions measure reported *ease of locomotion* – the ability to navigate effortlessly in the virtual world. The remaining seven questions measure the reported *presence*: the "sense of being" in a place or environment (e.g. a virtual environment) even when one is physically situated in another [18].

Usoh et. al [17] suggested to count the number of high answers (top 3 on a 7 point Likert scale) for the corresponding questions. Therefore presence is rated between 0 and 7 (number of possible high answers), whereas behavioral presence and ease of locomotion range from 0 to 3. Figure 2 shows the mean for each of the measures divided into the four groups. An analysis of variance for the UCL questionnaire showed highly significant results for the presence measure (Welch’s $F(3, 40.39) = 8.893, p < .001$).

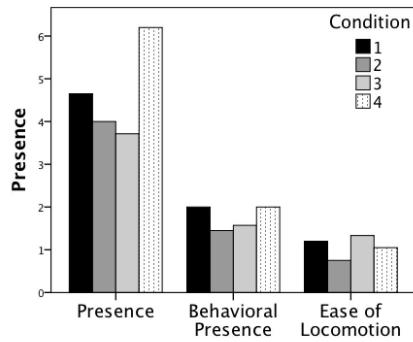


Fig. 2. Evaluation of the mean reported presence, behavioral presence and ease of locomotion. **Condition 1:** Graphics, **Condition 2:** Graphics & Audio, **Condition 3:** Graphics & Haptics, **Condition 4:** Graphics & Audio & Haptics

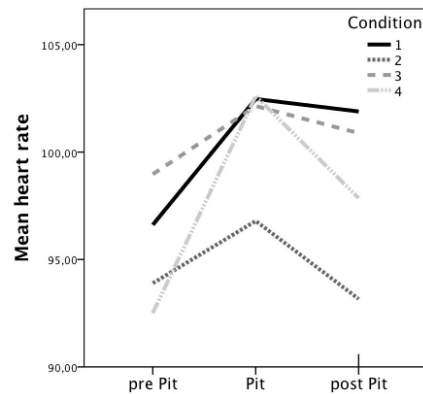


Fig. 3. Mean heart rate before, during, and after exposure to the pit room (measured in bpm). **Conditions 1, 2, 3,** and **4** see left.

Bonferroni’s post-hoc comparisons of the four conditions showed that participants in the second condition (Graphics & Sound) ($M = 4.00, SD = 2.03, CI[3.05, 4.95]$) as well as participants in the third condition (Graphics & Haptics) ($M = 3.71, SD = 1.31, CI[3.12, 4.31]$) rated the presence significantly lower than participants in Condition 4 (all modalities) ($M = 6.20, SD = 0.89,$

Table 1. Overview of the results from the presence related measures.

	Condition 1	Condition 2	Condition 3	Condition 4
Presented modalities	visual	visual auditory	visual haptic	visual auditory haptic
Presence	4.65 (2.16)	4.00** (2.02)	3.71** (1.31)	6.2** (1.91)
Behavioral Presence	2,00 (0.86)	1.45 (0.89)	1.57 (0.87)	2.00 (1.03)
Ease of Locomotion	1.2 (1.15)	0.75 (0.91)	1.33 (1.11)	1.05 (1.15)
Heart rate increase	5.85 (9.63)	2.90* (4.22)	3.17* (4.66)	10.10* (8.57)
Skin conductance increase	3.22 (2.51)	2.52 (1.66)	1.73 (1.04)	3.21 (3.11)

$CI[5.78, 6.62]$, $p < .001$. Condition 1 ($M = 4.65$, $SD = 2.16$, $CI[3.64, 5.66]$) was rated lower as well and is significant for $p < .05$. Behavioral Presence showed a similar non significant trend. Measures for Ease of Locomotion showed no significant differences across the four conditions.

As for the subjective measures of presence, the same analysis was performed for the measured heart rate and skin conductance. Due to equipment malfunction the physiologic data of six participants are missing. Figure 3 gives an overview of the measured heart rates for each group in each phase. 'Pre pit' is the whole training phase, 'pit' is the time from entering until leaving the pit room, and 'post pit' is measured for three minutes from leaving the pit room until the end of the experiment. In order to compare heart rates for training (pre pit) and pit room we calculated the increase for each group. The mean during the training phase and the mean while in the pit room were calculated and compared. An ANOVA showed significant results ($F(3, 70) = 4.1$, $p < .05$). Bonferroni post-hoc comparisons of the four groups indicate that heart rate increase in the fourth condition ($M = 10.1$, $SD = 8.57$, $CI[5.97, 14.22]$) was significantly higher than participants' heart rate increase in Condition 2 ($M = 2.89$, $SD = 4.22$, $CI[0.87, 4.93]$), $p < .05$ and 3 ($M = 3.17$, $SD = 4.66$, $CI[0.85, 5.48]$), $p < .05$. No significant difference but a similar trend was measured for skin conductance.

Table 1 gives an overview of these results. Mean and standard deviation (in parentheses) are given. For heart rate and skin conductance the increase from the training to the pit room is given. Significant values are marked with asterisks ($* = p < .05$; $** = p < .01$).

In addition correlation analyses were conducted for the presence related measures. The Score on the Immersion Tendency Questionnaire correlates with the reported presence ($r = .24$, $p < 0.05$) and behavioral presence ($r = -.29$, $p < 0.05$). Furthermore the score for reported *behavioral* presence correlates with gender (negative for male ($r = -.32$, $p < 0.01$)). The time spent playing com-

puter games did correlate with gender (for male $r = .25, p < 0.05$) but not with any results from the reported or observed presence related measures.

6 Discussion and Conclusion

Our results suggest that more presented modalities do not necessarily result in an increased perceived presence. In this study participants tended to rate presence lower when only presented with one additional (audio or haptic) cue. The recorded physiological data support this observation. When presented the full combination of modalities – the visual, the auditory, and the haptic – the perceived presence is significantly higher.

While this at first may seem counter-intuitive, it might be an uncanny valley effect in today’s virtual reality applications. As Masahiro Mori first stated for the robotics domain, if human replicas look and act almost, but not perfectly, like actual human beings, it causes a response of revulsion among human observers [12]. He called it a valley, corresponding to the valley on the graph of the comfort level of humans as a function of a robot’s human likeness. Brenton et al. described the same valley in the domain of virtual characters. If such characters are too close to a human but not perfect, people tend to dislike them [1]. A similar dip can be seen in our presence results (fig. 2), where Conditions 2 and 3 represent the valley.

One explanation could be that with all stimuli combined users’ expectations are better met and an increase in reported presence is observed. When presented only one additional cue, users may have expected more. Due to the overall availability of technology today, people are used to have a high amount of presented stimuli. For example, state-of-the-art computer games present at least very good auditory cues, and a lot of them tactile feedback as well. It may be not enough to present only a few sounds to make the world’s overall believability better. Our results suggests that the enhancement of virtual worlds with multi-modal stimuli, does not work in a simple additive fashion (or not anymore), like concluded by Dinh et al. [5].

Thus our future work will focus on how to improve users’ presence with multi-sensory stimuli. Doing so, further improvement of multi-modal feedback will be an important aspect, since it should be a relevant factor for the enhancement of immersion. There will be an effort to overcome the uncanny valley to increase the overall believability of and therefore make a step towards significantly improved immersion with additional multi-sensory stimuli. One approach could be the enrichment of virtual worlds with even more stimuli, for example heat or smell. Further the influence of “autonomy” factors, like the availability of an intelligent virtual agent as an interaction partner, seems worthwhile investigating as well.

7 Acknowledgments

This paper is a preprint version of an article published by Springer-Verlag. The original publication is available at http://link.springer.com/chapter/10.1007%2F978-3-642-39405-8_19.

References

1. H. Brenton, M. Gillies, D. Ballin, and D. Chatting. The uncanny valley: does it exist. In *In: 19th British HCI Group Annual Conference: workshop on human-animated character interaction*, 2005.
2. D. C. Cohen. Comparison of self-report and overt-behavioral procedures for assessing acrophobia. *Behavior Therapy*, 8(1):17 – 23, 1977.
3. C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: audio visual experience automatic virtual environment. *Commun. ACM*, 35(6):64–72, June 1992.
4. L. Deligiannidis and R. J. K. Jacob. The vr scooter: Wind and tactile feedback improve user performance. In *3DUI'06*, pages 143–150, 2006.
5. H. Q. Dinh, N. Walker, C. Song, A. Kobayashi, and L. F. Hodges. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings of the IEEE Virtual Reality, VR '99*, pages 222–228, Washington, DC, USA, 1999. IEEE Computer Society.
6. D. Fellner, J. Behr, and U. Bockholt. Instantreality - a framework for industrial augmented and virtual reality applications. In *The 2nd Sino-German Workshop "Virtual Reality & Augmented Reality in Industry"*.
7. E. Gibson and R. Walk. "the visual cliff". *Scientific American*, 202:64–71, 1960.
8. M. L. Heilig. Sensorama simulator, u.s.patent no.3050870, August 1962.
9. R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lillenthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.
10. M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr. Physiological measures of presence in stressful virtual environments. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques, SIGGRAPH '02*, pages 645–652, New York, NY, USA, 2002. ACM.
11. M. Meehan, S. Razzaque, B. Insko, M. Whitton, and J. Brooks, Frederick P. Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Applied Psychophysiology and Biofeedback*, 30:239–258, 2005.
12. M. Mori, K. MacDorman, and N. Kageki. The uncanny valley [from the field]. *Robotics Automation Magazine, IEEE*, 19(2):98 –100, June 2012.
13. R. Pausch, D. Proffitt, and G. Williams. Quantifying immersion in virtual reality. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH '97*, pages 13–18, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
14. P. Renner, T. Dankert, D. Schneider, N. Mattar, and T. Pfeiffer. Navigating and selecting in the virtual supermarket: Review and update of classic interaction techniques. In *Virtuelle und Erweiterte Realität: 7. Workshop der GI-Fachgruppe VR/AR*, pages 71–82. Shaker Verlag GmbH, 2010.

15. T. Sheridan. Further musings on the psychophysics of presence. In *Systems, Man, and Cybernetics, 1994. 'Humans, Information and Technology', 1994 IEEE International Conference on*, volume 2, pages 1073–1077 vol.2, Oct 1994.
16. M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 2(3):201–219, Sept. 1995.
17. M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoper. Virtual Environ.*, 9(5):497–503, Oct. 2000.
18. B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.