# When What You Hear is What You See: Presence and Auditory-Visual Integration in Virtual Environments

Pontus Larsson[1], Daniel Västfjäll[1,2], Pierre Olsson[3], Mendel Kleiner[1]
[1]Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden, [2]Department of Psychology, Göteborg University, Sweden, [3]Department of Building Design, Chalmers University of Technology, Göteborg, Sweden
{pontus.larsson@chalmers.se, daniel.vastfjall@psy.gu.se, pierre.olsson@arch.chalmers.se, mendel.kleiner@chalmers.se}

**Abstract**

*In this paper, it is hypothesized that consistency across modalities in terms of matching the visual space to the auditory space is important for the sense of presence. An experiment was carried out where thirty participants were exposed to four conditions having different degrees of auditory-visual consistency (one purely visual and three auditory-visual). A presence questionnaire was used after exposure to measure participants' sensations. Although participants rated the auditory-visual conditions as inducing significantly higher presence than the condition with only visual information, no differences in presence ratings between the three auditory-visual conditions were found. However, participants' rankings of their sensed presence in all conditions revealed that there might be such differences. Moreover, the results show that sound in general has a significant effect on VE users' sense of presence.*

*Keywords*--- **Presence, auralization, auditory-visual integration, virtual acoustics, multimodal VE**.

## 1. Introduction

The evaluation of high-end mediation technologies such as stereoscopic television, motion simulators and Virtual Reality (VR) has been a topic of interest during the past decade [1-3]. Employing a user-centered approach is often crucial in such evaluations to determine usability for VR interfaces and eventually reach market success [4]. One aspect of the perceptual and cognitive responses to mediation technologies, which has been of primary importance, is the user's sensation of being present inside the simulation (his or her "sense of presence") [4]. Presence can be seen as perceptual response to a mediated environment and should not be confused with immersion, which relates to the technology itself and its capabilities of enveloping the user [5]. The relation between presence and immersion has as of yet not been clearly established and one of the main future challenges is to find out how the various system parameters (visual, auditory, haptic etc.) correlates with the sense of presence of an average user.

Nonetheless, many of the mediation technologies that are currently in the focal point of presence research emphasize visual rendering of information [6]. It is most likely that technologies which also include advanced rendering of other types of sensory stimuli, such as audio and haptics, into a Virtual Environment (VE) is far superior in terms of generating a sense of presence simply because they represent real life to a higher extent [6]. Apart from enhancing the sense of presence, another advantage of systems including multi-modal information is that they may also increase users' performance of tasks and intuitive understanding of the system's functions.

However, little is known on exactly what types of sensory signals and combinations of sensory signals of a multimodal VE produce the highest presence responses. For example, it is as of yet an open and empirical question to which extent the sensory inputs, both across and within modalities, need to be consistent [5]. The lack of empirical data indicates the importance of conducting research into this area not only to establish new design criteria for multimodal VEs, but also to gain further theoretical understanding of the presence construct.

In this paper, we present an experiment where the relation between room acoustic rendering parameters and post-participation ratings of presence is explored. The aim of the study is twofold; first we attempt to replicate previous studies indicating that the addition of spatialization and auralization raises the sense of presence [7,6]. Second, we also aim at investigating whether physically congruent auditory-visual room information raises the sense of presence; whether it is also important that auditory and visual renderings both mediate the same room or if it is sufficient to add acoustical room information per se.

## 2. Background

As mentioned previously, presence research has mainly been focused on effects of visual stimuli. Although audition might be considered as our "secondary sense", less direct and subtler than vision, it is a fact that if the subtle auditory background of everyday life is suddenly removed (by e.g.

mufflers), we feel less connected to the world, less present [8]. Given also the fact that ears, in contrast to eyes, are always open and thus always susceptible to stimuli there should be no doubt that auditory percepts are profound components of the sensation of ‘being there’, no matter if ‘there’ is a real or a virtual place.

Auditory cues have received relatively little attention in the presence research area. In previous research it has been found that for a certain level of quality of the auditory stimulus added to a visual world, post-participation ratings of presence can be significantly increased [9]. An increase in quality or level of spatialization or auralization may also increase presence ratings [7,6].

However, research has also shown that enhancing spatial qualities of auditory cues may not always be beneficial [10]; the key to maximizing presence enhancement by auditory cues might have to involve distributing rendering and presentation effort in relation to the content of the virtual world. For example, the simulation of a ride in a rally car might benefit more from increased low frequency content or realistically rendered engine sound than realistically rendered spatial and room acoustic cues [10]. On the contrary, a strong sensation of being inside a large concert hall or a church most likely relies on high quality rendering of room acoustics properties such as early reflections and reverberation.

In close relation to this, the presence-enhancing effect of auditory stimuli may also be contingent on the congruity of the auditory and the visual virtual environment [6, 11, 12]. That is, presence is not likely to increase if we add a bicycle sound to a visual simulation of a rally car, or the sound of a living room to a visual simulation of a church.

What it all comes down to is that when designing sound and developing sound rendering systems for high-presence auditory-visual virtual environments, both spatial sound properties and sound source characteristics need to be considered in relation to the specific application and to the visual environment’s characteristics.

### 2.1. Presence, room information and auditory-visual synergy

Previous studies have shown that adding sound or other sensory stimuli to a visual VE may increase presence, but how should one design such stimuli in order to maximize the sense of presence? Stein & Meredith [13] state that “Integrated sensory inputs produce far richer experiences than would be predicted from their simple coexistence or the linear sum of their individual inputs”. Studies using basic stimuli have indeed shown that when sensory inputs are allowed to cooperate, auditory detection, visual choice reaction time, as well as sound localization may be improved [14-16]. For complex stimuli, such as realistic VEs, there is however a lack of research, which acknowledges Stein & Meredith´s concept of integrated sensory inputs. Nonetheless, it should be obvious that knowledge on how to produce such integrated sensory inputs could be of greatest importance to the design of VEs.

The idea of sensory integration when perceiving complex stimuli is elaborated by Larsson et. al. [17], where this concept is applied in the area of multi-modal perception of concert halls. In the model of room acoustic perception proposed by Larsson et. al, it is hypothesized that if information from the visual sense match aural information, the two modalities reinforce each other, are integrated and form a total evaluation. If this happens, sensory integration, perceptual fusion, or synergy is achieved, which means that information from auditory senses provides help for assigning meaning to information from visual senses and vice versa. On the other hand, if visual information is mismatched with aural information, ambiguity of interpretation arises and attention is directed on sorting out what is wrong. Information from different modalities may then be separated and processed separately rather than to be combined in an orderly fashion. In this case it is believed that listeners will rely on visual information (visual dominance) rather than auditory information when forming a total evaluation.

In the model by Larsson et. al. [17], the occurrence of perceptual fusion or integration and modality mismatch effects is also contingent on the listener's cognitive, evaluative and affective reactions. That is, experience, expectations, and attitudes towards the room, music, performer, and total situation will influence the rise of synergy or modality mismatch.

Moreover, Larsson et. al. argues that synergy effects may arise in both real and virtual environments, while they are less likely to occur in virtual environments. This is partly due to the limitations of the currently available VR-technology and the fact that stimuli for different modalities are often independently rendered. However the potential lack of synergy effects in VEs may also be due to the fact that little is known on exactly how the stimuli for different modalities should be designed in order to obtain synergy effects.

Is then synergy a prerequisite for high presence? It seems likely that if both vision and audition tells us that we ‘are there’, presence should dramatically increase if, and only if, ‘there’ is perceptually the same place visually and auditorily. If not, a reasonable assumption would be that we feel as being present in either the visual world or in the auditory world, depending on the user’s field of dominance, the preferred sensory modality [5]. For example, if a visually dominant person is exposed to a VE where the auditory VE does not correspond to the visual VE, the auditory cues do not provide any useful information in the person’s internal representation of the virtual world. The person can then only be present in the visual world. The result from presenting such conflicting auditory/visual environments may simply also be confusion.

## 3. Experiment

In the current experiment, we study the effects of physically mismatched and matched aural and visual inputs on presence by using different combinations of visualizations and auralizations. The hypothesis is that when matched auditory and visual stimuli is presented, perceptual fusion will occur which in turn will lead to a significant increase in sensed presence. Our definition of ‘matched’ is in this case that auditory and visual stimuli have been created using the same physical model.

### 3.1. Stimuli and design

The stimuli used in current experiment were based on a digital model of Örgryte New Church in Göteborg, Sweden (see Figure 1).

**Figure 1 Screenshot of the visual digital model used in the experiment**

The volume of the church is approximately 8010 $m^3$. The software dVision dvmockup was used to present the digital model visually and to enable real-time interaction. The acoustic simulation was rendered with the auralization software CATT-Acoustic in combination with a Huron PCI real-time convolution workstation. Acoustic simulation input consisted of organ recordings made at close distance in the actual church (see Figure 2). Eight different microphone positions were used, three outside the organ case and five located on various places inside the organ case. This virtually anechoically recorded sound was then used as input signal to the acoustic simulation.

**Figure 2 One of the microphone positions used in the near-field recordings of the organ: Inside the main case**

The visual and auditory simulations were spatially synchronized, i.e. the user’s position and orientation in the visual simulation was continuously transferred to the Huron workstation. Each participant experienced four different experimental conditions corresponding to different levels of auditory rendering used in combination with the visual stimuli; 1) No sound, 2) Non-auralized, anechoic sound 3) Auralized “mismatched “sound, 4) Auralized “matched” sound. In the second condition, no processing was applied to the anechoic sound before sent to loudspeakers. In the third condition, mismatched auralized sound condition, the model used to render the acoustics was a generic, shoebox-shaped room, having a volume of 2000 $m^2$. In the fourth condition, the auditory scene was rendered using the model of the church (having a volume 8010$m^3$). Half of the participants were exposed to the conditions in order 1, 2, 3, 4 while the other half were exposed to the conditions in reversed order (4, 3, 2, 1).

### 3.2. Instrumentation

For all conditions a CAVE-like immersive display system (TAN VR-CUBE) was used to display the virtual environment. The VR-CUBE is a 3x3x3 m cubic room with 3D graphics display on all four walls and on the floor. The graphics is rendered by a SGI Onyx2 InfiniteReality2 system with 8 MIPS R10000 CPUs (running at 195 MHz), 2 GB RAM and three graphic pipes and displayed on the five screens by Electrohome Marquee 8500/p43 projectors. Time-multiplexed stereoscopic presentation is enabled by Stereographics LCD shutter glasses. Furthermore, a Polhemus 3space Fastrak Longranger electromagnetic tracking system keeps track of the viewer’s head position so that the perspective is projected correctly on the display. The VR-CUBE’s audio system is based on a Lake Huron PCI computer with 12 DSPs for real-time auralization and a standard PC with the Gigasampler software for playback of anechoic sound files. The current experiment used the Ambisonic format for sound presentation via eight active monitor loudspeakers mounted in each corner of the VR-CUBE. The loudspeakers located behind screens (i.e. the four loudspeakers standing on the floor) were individually equalized to compensate for the sound transmission loss through the screens. A Polhemus 3D joystick was used as input device, which allowed the participants to move either forwards or backwards by pressing two different buttons on the joystick. By changing head orientation, participants could change their direction of travel.

### 3.3. Participants

Thirty undergraduates and graduates, four female, from Chalmers University of Technology participated on voluntary basis. Their mean age was 32 years (SD. 9.0). A within-group design was applied so that all participants performed the experiment in four different conditions in a sequence.

### 3.4. Task

Participants' task was to find and navigate through five numbered 1x1 meter cubes. The intention of this task was to ensure that all participants navigated around in the church in approximately same manner. Participants were instructed to navigate to and fly through the cubes in numerical order and to return to the starting point after finding all five cubes.

### 3.5. Measures

Measures of quality of as well as reactions to VEs were taken from the Swedish viewer-user presence questionnaire (SVUP) [18]. The SVUP questionnaire comprises in total 150 items covering quality evaluations, attitudes, presence, and realism, and information from different modalities as well as simulation sickness items. In this experiment, the short version of SVUP (SVUP-short) was used which comprises 18 items covering VE interaction, presence, awareness of external factors, sound quality, enjoyment, and simulation sickness items were selected. A standard 7-point scale with the verbal endpoints "not at all" and "very much" was used. Below, the 18 SVUP-short items are shown (translated from Swedish).

- How natural was the interaction with the environment?
- How involved were you in the experience?
- To what extent did you feel that you were present in the virtual environment?
- To what extent did you think that the things you did and saw happened naturally and without much mental effort?
- To what extent were you aware of things happening around you, outside the virtual environment?
- To what extent did you feel disoriented or confused in the virtual environment?
- To what extent did you focus your attention on the situation, rather then other things?
- To what extent did you think it was enjoyable to interact in the virtual environment?
- To what extent did you find the virtual environment fascinating?
- To what extent were you able to identify sounds?
- To what extent were you able to localize sounds?
- To what extent did you think that the sound contributed to the overall realism?
- I felt nauseous
- My eyes felt strained
- I felt dizzy
- I had a headache
- I had problems concentrating
- I felt unpleasant

A 0-100 Presence scale was also used with the verbal endpoints "completely unreal" and "just like in real life". In addition, 15 participants from the reverse order group (see "stimuli and design" above) were asked to rank their sensations of presence in the four experimental conditions.

### 3.6. Procedure

Participants arrived individually at the laboratory and were directed to the VR-CUBE. A male experimenter instructed participants on the task, to find and fly through the five numbered cubes, and on the use of the joystick and the shutter glasses. Participants then completed the task in the VE. After completion of the task the participants were handed a single page questionnaire comprising the 18 SVUP-short items and the 0-100 presence scale. This procedure was repeated for all experimental conditions (no sound, sound, auralized "mismatched" sound, auralized "matched" sound). After completing the four experimental condition tasks and corresponding questionnaires, participants were asked to fill in a single page questionnaire comprising 11 questions concerning age, gender and previous experiences of psychological tests and VR-systems. 15 participants were also asked to rank the four conditions in terms of the sense of presence experienced. Participants were then debriefed and thanked for their participation.

### 3.7. Results

The item ratings were submitted to separate independent t-tests to determine between conditions differences. Bonferroni's method was used to adjust for multiple comparisons.

The first thing to note is that the analysis show that all three sound conditions were rated as having more natural interaction than the no sound condition.

| | CONDITIONS | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1: No sound | | 2: Sound | | 3: Mismatched auraliz. | | 4: Matched | | |
| **ITEM** | M | SD | M | SD | M | SD | M | SD | p |
| Natural | 4.28 | 0.22 | 4.80 | 0.27 | | | | | < 0.05 |
| | " | " | | | 4.86 | 0.24 | | | < 0.01 |
| | " | " | | | | | 4.78 | 0.27 | < 0.05 |
| Ext. Awareness | 3.43 | 0.29 | 2.65 | 0.30 | | | | | < 0.05 |
| | " | " | | | *2.80* | *0.32* | | | *ns (p= 0.25)* |
| | " | " | | | | | 2.61 | 0.31 | < 0.05 |
| Disoriented | 2.64 | 0.28 | 2.02 | 0.18 | | | | | < 0.05 |
| | " | " | | | *2.12* | *0.20* | | | *ns (p=0.31)* |
| | " | " | | | | | *1.85* | *0.20* | *ns (p=0.076)* |
| Involved | *4.63* | *0.21* | *5.06* | *0.27* | | | | | *ns (p=0.15)* |
| | " | " | | | 5.17 | 0.24 | | | <0.05 |
| | " | " | | | | | *5.17* | *0.20* | *ns (p=0.19)* |
| Presence 1-7 | 4.45 | 0.21 | 5.21 | 0.26 | | | | | <0.01 |
| | " | " | | | *4.99* | *0.23* | | | *ns (p=0.087)* |
| | " | " | | | | | *5.03* | *0.20* | *ns (p=0.098)* |
| Presence 0-100 | 49.15 | 3.99 | 57.57 | 4.34 | | | | | <0.01 |
| | " | " | | | 58.64 | 3.92 | | | <0.01 |
| | " | " | | | | | 61.20 | 3.93 | <0.01 |
| Presence index (Cronbach's alpha = 0.75) | 4.47 | 0.19 | 5.02 | 0.25 | | | | | <0.05 |
| | " | " | | | 5.01 | 0.20 | | | <0.01 |
| | " | " | | | | | 5.00 | 0.19 | <0.05 |
| Sound add to realism | | | *4.77* | *0.37* | *5.17* | *0.32* | | | *ns (p=0.2)* |
| | | | " | " | | | 5.80 | 0.26 | <0.01 |
| Localize sound | | | *3.44* | *0.39* | *3.85* | *0.36* | | | *ns (p=0.49)* |
| | | | " | " | | | 4.44 | 0.32 | <0.05 |

**Table 1 Summary of significant differences between sound conditions. The table shows differences between the no sound condition and the three sound conditions except for the last two items, where the sound condition is compared with the mismatched and matched auralization conditions.**

Concerning the external awareness scale, conditions 2 and 4 (sound and matched auralized sound) differed significantly ($p< 0.05$) from condition 1 (no sound), M=2.65 and M=2.61 vs. M=3.43 indicating that participants were less aware of what happened outside the VE in the two sound conditions. Although the difference is not significant ($p=0.25$), one might note a slight trend for condition three to induce less external awareness in participants (M=2.80).

For the disoriented scale, participants were less disoriented in the 2nd condition (sound, no auralization) compared to the no sound condition (M=2.02 vs. 2.64, $p<0.05$).

For the 1-7 presence scale, the only significant difference found was between conditions 1 (no sound) and 2 (sound, no auralization) where participants rated condition 2 as inducing more presence than condition 1 (M=5.21 vs. M=4.26, $p<0.01$). However, the analysis showed that ratings on the 0-100 magnitude estimation presence scale in all sound conditions (conditions 2-4) differed significantly from ratings in the no sound condition (M=57.6, M=58.6, and M=61.2 vs. M= 49.2; $p<0.01$).

On basis of previous principal component analysis (PCA) of experimental results obtained with the SVUP questionnaire [6], three presence items (natural, present and involved) were also averaged into a presence index[1]. Similarly, the enjoyment (two items), external awareness (two items) and simulation sickness (six items) were averaged into separate index variables. The averaged index ratings were then submitted to separate independent t-tests to determine between-conditions differences. The results from the analysis of presence index yielded similar results as the magnitude estimation presence scale; presence index ratings were significantly higher in the sound conditions compared to the no sound condition. No differences between conditions could be observed in the three other indices.

Furthermore, analysis of the sound items in the three sound conditions showed that participants rated condition 4 (matched auralization) as having a sound that added more to the overall realism than condition 2 (sound, no auralization) ($M=5.80$ vs. $M=4.87$, $p<0.01$). Similar results were obtained for the localization item; participants found that the sound was more easily localizable in condition 4 compared to condition 2 ($M=4.44$ vs. 3.44, $p<0.05$).

In sum, participants rated their sense of presence higher in the sound conditions compared to the no sound conditions. Moreover, participants ratings also indicate that the interaction was more natural and that the participants were less disoriented in the sound conditions compared to the no sound condition. Concerning the sound conditions, participants rated the sound as adding more to the overall realism and as being more easily localizable in condition 4 compared to condition 2. The results are summarized in Table 1.

As can be seen in the results of the analysis above, participants in general only discriminated between the no sound and the three sound conditions. That is, there was no interaction between different levels of auralization and e.g. presence, which goes counter to the hypothesis and previous findings [6]. In order to analyze this in more detail, Pearson's correlation coefficient between participants' ratings of "sound's contribution to the overall realism" and presence was calculated.

The results of the correlation analysis showed that presence ratings (given by the averaged presence index) are positively correlated to the sound contribution item for all three conditions (0.750, 0.660 and 0.536, $p<0.01$), indicating that participants who found the sound to be more contributing to the overall realism also gave higher presence ratings. Similar results were found when correlating the magnitude estimation presence scale and the 1-7 presence scale with the sound contribution item, although significance was not reached for the fourth condition in these analyses.

Furthermore, after examining in more detail the distribution of the presence ratings (both the 1-7 and the magnitude estimation scale) a difference between the two groups of participants was revealed; The presence ratings by participants in the second group (who were exposed to the experimental conditions in order 4,3,2,1) tend to be distributed over a wider range of the scale compared to the ratings of the first group (who were exposed to the experimental conditions in order 1,2,3,4). The reason for this might be that participants in the first group gave reasonably high ratings of presence already in the first condition and thus had little room to give higher ratings in the next three conditions (i.e. a ceiling effect).

This order effect may be observed from the second group (reverse order) participants who were asked to rank the different conditions by how much presence they had experienced in each condition (Figure 3).

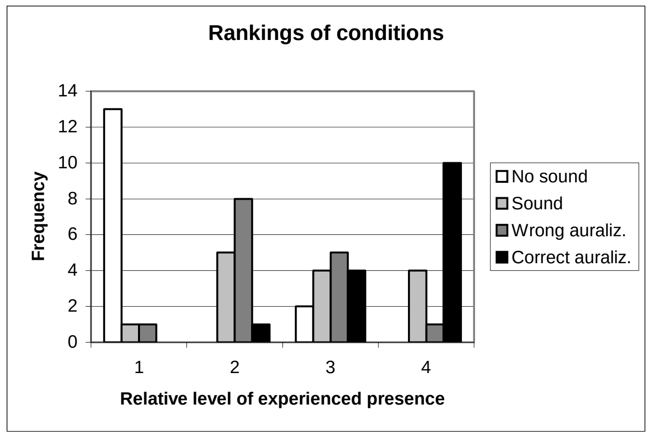


**Figure 3 Group 2's rankings of their sense of presence in the different conditions.**

As can be seen, the no sound condition is clearly rated as being the least presence inducing condition. This is substantiated by an analysis where we test deviation from chance (in Figure 3, the expected frequency for each category is 3.8 if the ratings would have been random), $chi^2$ (1)=31.1, $p<0.001$. Moreover, in line with our hypothesis, the correct auralization condition is rated as being the most presence-inducing condition, $chi^2(1)= 16.2$, $p<0.001$. There is no clear tendency for the two other conditions.

## 4. Discussion

It was hypothesized that consistency across modalities, in terms of physical renderings of the auditory and visual spatial properties of a room, is a prerequisite for high presence. From the results presented in the previous section, it can be seen the main differences in SVUP ratings were found when comparing the non-sound condition with the three different sound conditions. In general, the no-sound condition generated lower presence ratings and higher sense

[1] One item in the original subscale (mental effort) was, due to low item-scale correlation, excluded from this index.

of external awareness. The predicted differences in presence ratings between the various sound conditions were thus not obtained in the current experiment, although participants' post-ranking of the conditions revealed that there might have been a preference for the correct auralization condition. Furthermore, the finding that the "sound contribution to overall realism"-item was positively correlated with presence ratings to some extent supports the idea of field dominance [5]; that only some participants fully perceived the sound as contributing to the feeling of presence.

There are several possible explanations to these findings. One may be that, although special care was put into the design and presentation of the visual stimulus, participants did not get the correct visual spatial impression of the church interior, i.e. the one corresponding to the correct auralization condition of sound.

This in turn may be due to the fact that, although the model used in the current experiment was highly detailed in terms of number of polygons, only a limited number of textures were applied to the model's surfaces. Texture is known to be a strong cue to distance [19], and the lack of a continuous textured surface may cause inaccurate distance judgments in the real world [20]. It is not entirely clear to which extent or in what way textures influence distance perception in VEs [21], but it might have been the case that the low number of textures in the current model led to an underestimation of distances and thus also an underestimation of the visually perceived room size.

Moreover, recent studies indicate that eyepoint height affect perception of distances and dimension in VEs, in that eyepoint height is negatively correlated with distance perception [22]. That is, the higher the eyepoint height, the lower the estimations of distances. It therefore seems necessary to adapt and fix the eyepoint height to the user in order for the visual spatial percept to be correct. In the current application, the observer movements were not restricted in any way, which may have given participants the sensation that the church was smaller than what it actually is.

Nonetheless, if incorrect visual spatial perception was the case in the current experiment, it seems likely that participants should have rated both auralized versions of the church as inducing a higher sense of presence than the non-auralized version, given the results by Larsson et. al. [6]. Another possible explanation is that the audio system used in the current study simply did not provide high enough display quality to be at all comparable to the visual display fidelity. The Ambisonic reproduction system is known to be less distinct in terms of e.g. localization compared to binaural sound [23] and in the current setup, the audio display quality is also limited due to the placement of the loudspeakers. By having this arrangement, it may have been the case that the subjects were not able to detect the nuances of the sound but were more relying on the visual stimulus.

Yet another explanation to the results is that it is perhaps not the spatial characteristics that are the most salient determinants of auditory induced (or enhanced) presence for the current scenario. Previous research has shown that the sound source identity and the expectancy of what sound sources to be present in a certain situation are likely to affect the sensation of presence [12]. In our case, hearing the sound of an organ is clearly and naturally associated with visually experiencing a church interior, and perhaps more important than hearing the spatial acoustic properties of the church. Therefore, it might have been the case that the effect of the strong sound source to visual environment identity overruled any effects of the added or improved auralization. This may imply that auditory-visual matching in some situations is a matter of providing the right sound source and not (just) the right acoustics.

To conclude, it is clear that further experiments have to be conducted in order to provide solid guidelines for the design of auditory-visually matched and perceptually optimized VEs. Such experiments could either further explore the relation between auditory and visual spatial cues or investigate the matching between sound source content cues and visual cues. Perhaps even more important is to establish the interaction effects between spatial cues and source content. That is, to investigate how the attention is distributed to either source identity or to source spatialization for different types of situations, scenarios or content, and for users with different preferences of modality. Such research will not only be instrumental in the understanding of auditory-visual integration and presence, but is probably also beneficial for the development of related technology and how computational resources should be allocated at any given moment.

## Acknowledgements

The work presented in this paper was supported by the EU project ORSEV and the EU FET Presence 1&2 projects POEMS (IST-2001-39223) and PRESENCCIA (contract no. 27731).

## References

[1] W. A. Ijsselsteijn. Presence in depth. Eindhoven: Technische Universiteit Eindhoven. 2004.

[2] P. Larsson, D. Västfjäll, M. Kleiner. Perception of self-motion and presence in auditory virtual environments. In: Proceedings of the Presence Seventh Annual International Workshop, Valencia, Spain, October 13-15: 252-258, 2004.

[3] D. A. Bowman, D. Raja. A method for quantifying the benefits of immersion using the CAVE. Presence-Connect. 4(7). 2004.

[4] M. Lombard, T. Ditton. At the heart of it all: The concept of presence. Journal of Computer Mediated Communication. 3(2). 1997.

[5] M. Slater. A note on presence terminology. Presence-Connect. 3(1). 2003.

[6] P. Larsson, D. Västfjäll, M. Kleiner. Spatial auditory cues and presence in virtual environments. Submitted for publication.

[7] C. Hendrix, W. Barfield. The sense of presence within auditory virtual environments. Presence: Teleoperators and Virtual Environments. 5(3): 290-301. 1996.

[8] R. H. Gilkey, J. M. Weisenberger. The sense of presence for the suddenly deafened adult: Implications for Virtual Environments. Presence: Teleoperators and Virtual Environments. 4(4): 357-363. 1995.

[9] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, A. Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In: Proceedings of IEEE Virtual Reality, Houston, TX, March 13-17: 222 – 228. 1999.

[10] J. Freeman, J. Lessiter. Here there & everywhere: The effects of multi-channel audio on presence. In: Proceedings of The Seventh International Conference on Audio Display, Helsinki, Finland, July 29 - August 1: 231-234. 2001.

[11] K. Ozawa, S. Ohtake, Y. Suzuki, T. Sone. Effects of visual information on auditory presence. Acoustical Science and Technology. 24(2): 97-99. 2003.

[12] P. Chueng, P. Marsden, P. Designing auditory spaces: The role of expectation. In: Proceedings of 10th International Conference on Human Computer Interaction, Crete, Greece, June 22-27: 616-620. 2003.

[13] B. E. Stein, M. A. Meredith. The merging of the senses. Cambridge, MA: MIT press. 1997.

[14] S. Komiyama. Subjective evaluation of angular displacement between picture and sound direction for HDTV systems. Journal of Audio Engineering Society. 37: 210-214. 1989.

[15] S. D. Lipscomb. Cross-modal integration: Synchronization of auditory and visual components in simple and complex media. In: Proceedings of Forum Acusticum, Berlin, Germany, March 14-19. 1999

[16] J. Simon, J. Craft. Effects of an irrelevant auditory stimulus on visual choice reaction time. Journal of Experimental Psychology. 86: 272-274. 1970.

[17] P. Larsson, D. Västfjäll, M. Kleiner. Auditory-visual interaction in concert halls. Submitted for publication.

[18] P. Larsson, D. Västfjäll, M. Kleiner. The actor-observer effect in virtual reality presentations. CyberPsychology & Behavior. 4(2): 239-246. 2001.

[19] J. J. Gibson. The ecological approach to visual perception. Hillsdale, NJ: Lawrence Erlbaum Associates. 1979.

[20] M. J. Sinai, T. L. Ooi, Z. J. He. Terrain influences the accurate judgement of distance. Nature. 95: 497-500. 1998.

[21] M. J. Sinai, W. K. Krebs, R. P. Darken, J. H. Rowland, J. S. McCarley. Egocentric distance perception in a virtual environment using a perceptual matching task. In: Proceedings of the 43rd Annual Meeting Human Factors and Ergonomics Society, Houston, TX, September 27 - October 1: 1256-1260. 1999.

[22] G. Franz, M. von der Heyde, H. H. Bülthoff. SceneGen - Automated 3D scene generation for psychophysical experiments. Presented at the 6th Tübinger Wahrnehmungskonferenz, Tübingen, Germany, February 21-23. 2003.

[23] A. Farina, E. Ugolotti. Subjective comparison between stereo dipole and 3D Ambisonics surround systems for automotive applications. In: Proceedings of the 16th Audio Engineering Society Conference, Rovaniemi, Finland, April 10-12: 16-048. 1999.