Increasing the Motion of Users in Photo-realistic Virtual Environments by Utilising Auditory Rendering of the Environment and Ego-motion

Rolf Nordahl

Medialogy, Aalborg University Copenhagen
2750 Ballerup, Denmark
{rn@media.aau.dk}

Abstract

An occurring problem of the image-based-rendering technology for Virtual Environments has been that subjects in general showed very little movement of head and body. Our hypothesis is that the movement rate could be enhanced by introducing the auditory modality. In the study described in this paper, 126 subjects participated in a between-subjects experiment involving six different experimental conditions, including both uni- and bi-modal stimuli (auditory and visual). The aim of the study was to investigate the influence of auditory rendering in stimulating and enhancing subjects' motion in virtual reality.

The auditory stimuli consisted of several combinations of auditory feedback, including static sound sources as well as self-induced sounds. Results show that motion in virtual reality is significantly enhanced when moving sound sources and sound of ego-motion are rendered in the environment.

1. Introduction

Although sound is one of the fundamental modalities in the human perceptual system, it is still a largely undiscovered area to researchers and practitioners of Virtual Reality. While research has provided insights into aspects of how the multimodal nature of the human may consist, many questions still remain in how one can utilize e.g. audio-visual phenomena when building new media products.

In previous research, we investigated if self-sound enhanced sense of presence in virtual reality [1]. These environments were based on the Image Based Rendering technique [13]. While the users were exposed to photorealistic virtual environments, where they could explore the environment within the constraints of the Region of Exploration (with a radius of 60 cm), no temporal information was delivered. Such virtual environments could be suggested to be a good example of the current state of Presence and VR research, where the feeling of Presence is considered by some to be mostly linked to the spatial domain [14]. However, we believe that such feeling of Presence may be heightened by adding temporal or spatio-temporal information, e.g., created by auditory feedback.

In order to examine this assumption we previously designed parametric sound effect [2], controlled by the motion of a user in the virtual environment. This was done using sound synthesis by physical models. Such physical models described the sonic impact between the user wearing a shoe and the different surfaces that were designed to accommodate the visual stimulus. These models were driven by a footstep controller which the user was asked to wear before entering the VR experience. The addition of sound effects created some interesting dynamic variations in the otherwise static environment, and the experiments showed how interactive sounds enhanced the sense of presence in VR [1].

In the previous simulated virtual environment no other sounds but the one generated by the footsteps controller and synthesizer were present. This was done because the visual feedback proposed to the subjects was an empty technical museum in Prague, where no sounds other than the noise of a fan could be heard. Drawing a parallel to current film sound theory we therefore assumed that introducing other sounds, would act as non-diegetic [12], i.e., sounds which perform as something outside the "reality" of the virtual world, and therefore conflict with the overall goal of creating a close-to-reality experience.

We observed that subjects appreciated the quality of the images reproduced by image-based rendering techniques. However, they were not very stimulated to visit the environment, since no dynamic events were happening. We are therefore interested in investigating if by adding auditory feedback, and therefore a temporal dimension, it is possible to enhance the interest of subjects to such a degree that they actively investigate and explore the environment.

In particular, we extended our previous research by adding environmental sounds as well as simulation of moving sound sources to the virtual reality experience. The sources of such sounds, which might be located out of the subject's field of view, are diegetic, i.e., belonging to the constructed reality and therefore sustain the experience. Our goal is to understand if enhancing the environment with dynamic sounds increases the motion of subjects in virtual environments. Our hope is that sound creates dynamic variations which make the environment more interesting to visit and investigate.

In order to achieve this goal, we decided to monitor the motion of subjects in the virtual environment. We

believe that a higher degree of motion represents a higher degree of engagement and interest in exploration of the simulation.

2. Experiment design

2.1 PC and peripherals for Visual Delivery

In the following section the hardware and software configuration used in the experiments is described. The visual stimulus was provided by a standard PC running Suse Linux 10. This computer was running the BENOGO software using the REX disc "Prague Botanical Garden".

The Head-Mounted-Display used was a VRLogic V8[2]. It features Dual 1.3" diagonal Active Matrix Liquid Crystal Displays with resolution per eye: ((640x3)x480), (921,600 color elements) equivalent to 307,200 triads. Furthermore the HMD provides a field of view of 60° diagonal.

The tracker used was a Polhemus IsoTrak II[3]. It provides a latency of 20 milliseconds with a refresh rate of 60 Hz.

2.2 PC and peripherals for audio delivery

The audio system was created using a standard PC running MS Windows XP SP 2. All sound was run through Max/MSP 4.5.1[4] and as output module a Fireface 800 from RME[5] was used. Sound was delivered by eight Dynaudio BM5A speakers[6].

2.3 Audio-visual setup

In the laboratory eight speakers were positioned in a parallelepipedal configuration. Current commercially available sound delivery methods are based on sound reproduction in the horizontal plane. However we decided to deliver sounds in eight speakers and thereby implementing full 3D capabilities. By using this methodology we were allowed to position both static sound elements as well as dynamic sound sources linked to the position of the subject. Moreover we were able to maintain a similar configuration to other virtual reality facilities such as CAVEs, where eight channels surround is presently implemented. This is the reason why 8-

channels sound rendering was chosen compared to e.g. binaural rendering.

As described two computers were installed in the laboratory, one running the visual feedback described in the following section, and one running the auditory feedback. A Polhemus tracker, attached to the head mounted display, was connected to the computer running the visual display, and allowed to track the position and orientation of the user in 3D. The computer running the visual display was connected to the computer running the auditory display by TCP/IP. Connected to the sound computer there was the interface RME Fireface 800 which allowed delivering sound to the eight channels, and the wireless shoe controller. The mentioned controller, developed specifically for these experiments [10], allowed detecting the footsteps of the subjects and mapping these to the real-time sound synthesis engine. The different hardware components are connected together as shown in Figure 1.
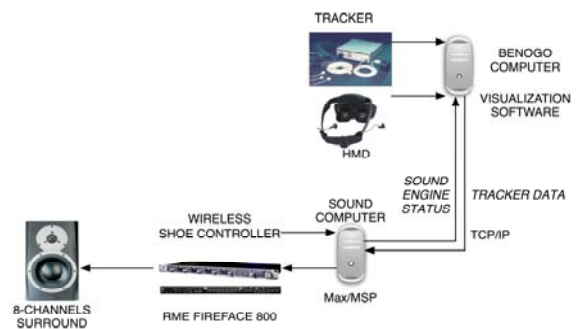

Figure 1. Technical configuration to connect the different hardware components necessary for the experiment

3. Audio-Visual feedback

3.1 Visual feedback

The visual feedback used in these experiments was created under the BENOGO[7] project.

The idea behind this project is the creation of photorealistic visual environments obtained by taking pictures of a specific location at different angles, and building a reconstruction of the same place at the computer using image based rendering techniques.

---

[2] http://www.vrlogic.com/html/virtualresearch/v8.html (accessed January 17th, 2006)

[3] http://www.inition.co.uk/inition/pdf/isotrak.pdf (accessed January 17th, 2006)

[4] http://www.cycling74.com/ (accessed January 17th, 2006)

[5] http://www.rmeaudio.com/english/firewire/ff800.htm#TECHNICAL (accessed January 17th, 2006)

[6] http://www.dynaudioacoustics.com/Default.asp?Id=3680&AjrNws=663&AjrNwsPg=1(accessed January 17th, 2006)

[7] www.benogo.dk

Figure 2: A view of the botanical garden in Prague reconstructed using image based rendering techniques.

One of the peculiarities of this approach is the fact that no moving objects have to be present in the environment when the pictures are taken, since this would affect the visual reconstruction.

This also implies that the reconstructed scenarios do not vary over time, which means that one could be concerned with that the exposure to the environment becomes tedious and uninteresting for the users to explore. As such we regard the exploration of auditory feedback as a good way to cope with these limitations, as explained in the following section.

3.2 Auditory feedback

Different levels of auditory rendering were implemented, in order to test their effect on stimulating motion of subjects in virtual reality.

A soundscape accompanying the visualization of the Prague botanical garden shown in Figure 2 was designed. Such soundscape contained sounds of birds, water dropping sounds, and other kinds of environmental sounds typical of a botanical garden. The soundscape was delivered using 8-channel surround sound speakers in a damped laboratory via Dynaudio BM5A loudspeakers.

Additionally, moving sound sources such as mosquitoes flying around the room were added, to enhance the sense of immersion in the soundscape.

Furthermore, ego-sounds of the subjects walking around the environment were synthesized in real-time.

Such sounds were created by having the users wear a pair of pressure sensitive sandals, connecting to a PC via a wireless interface. Walking around the room, subjects generated in real-time walking sounds synthesized using modal synthesis.

To summarize, four kinds of auditory stimuli were provided to the subjects:

1. "Static" soundscape, reproduced at max. peak of 58dB, measured c-weighted with slow response. This soundscape was delivered through the 8-channels system.
2. Dynamic soundscape with moving sound sources, developed using the VBAP algorithm [9], reproduced at max. peak of 58dB, measured c-weighted with slow response.
3. Auditory simulation of ego-motion, reproduced at 54 dB. (This has been recognised as the proper output level as described in [1])
4. A piece of classic music [11], reproduced at max. peak of 58dB, measured c-weighted with slow response.

Figure 3: A subject during the experiment.

4. Test description

126 subjects took part to the experiment. All subjects reported normal hearing and visual conditions. Figure 3 shows one of the subjects participating to the experiment.

Before entering the room, subjects were asked to wear a head mounted display and the sandals enhanced with sensors. Subjects were not informed about the purpose of the sensors-equipped footwear. Before starting the experimental session the subjects were told that they would enter a photo-realistic environment, where they could move around if they so wished. Furthermore, they were told that afterwards they would

have to fill out a questionnaire, where several questions would be focused on what they remember having experienced. No further guidance was given. Each subject was exposed to one of the six conditions shown in Table 1 for 3 minutes. The experimental conditions are described in the following section.

4.1 Experimental Conditions

The experiment was performed as a between-subjects study including the following six conditions:

1. Visual only. This condition had only uni-modal (visual) input.

2. Visual with footstep sounds. In this condition, the subjects had bi-modal perceptual input (audio-visual) comparable to our earlier research [1].

3. Visual with full sound. This condition implies that subjects were treated with full perceptual visual and audio input. This condition included static sound design, 3D sound (using VBAP algorithm [9]) as well as rendering sounds from ego-motion (the subjects triggered sounds via their footsteps).

4. Visual with full sequenced sound. This condition was strongly related to condition 3. However, it was run in three stages: the condition started with bi-modal perceptual input (audio-visual) with static sound design. After 20 seconds, the rendering of the sounds from ego-motion was introduced. After 40 seconds the 3D sound started (in this case the sound of a mosquito).

5. Visual with sound + 3D sound. This condition introduced bi-modal (audio-visual) stimuli to the subjects in the form of static sound design and the inclusion of 3D sound (the VBAP algorithm using the sound of a mosquito as sound source). In this condition no rendering of ego-motion was conducted.

6. Visual with music. In this condition the subjects were introduced to bi-modal stimuli (audio and visual) with the sound being a piece of classical music [11]. This condition was used as a control condition, to ascertain that it was not sound in general that may influence the in- or decreases in motion. Furthermore it enabled us to deduce if the results recorded from other conditions were valid. From this it should be possible to deduct how the specific variable sound design from the other experimental conditions affects the subjects.

Table 1 summarizes the different experimental conditions together with the mean and standard deviation of the age of the participants. The first column of Table 1 represents the name of the condition, in the same order as described above, while the second column outlines the auditory feedback of each condition as described in Section 3.2. As an example, the condition "full" represents the situation in which subjects were exposed to auditory feedback 1,2 and 3, which means a static soundscape, a dynamic soundscape where moving sound sources were introduced, and the sound produced by their footsteps while they were walking around the virtual environment.

CONDITION	AUDITORY STIMULI	NUMBER SUBJECT	MEAN (AGE)	ST.D. (AGE)
Visual	None	21	25.6	4.13
Visuals w. foot	3	21	25.7	3.75
Full	1 + 2 +3	21	25	4.34
Full seq	1 + 2 + 3	21	22.8	2.58
Sound + 3D	1+2	21	22.9	2.5
Music	4	21	28	8.1

Table 1: Description of the seven different conditions to which subjects were exposed during the experiments. The number in the second column refers to the auditory feedback described in Section 3.2.

It is worth reminding that in the fourth condition called full sequenced (abbreviated with "full. Seq."), subjects were exposed to the same auditory stimuli as in the first condition. The difference consists in the order of appearance of such stimuli. For example, the ego-sound generated by the footsteps was introduced only 20 seconds after the subjects started the experiment, whereas in the full condition the ego-sound was present from the beginning of the experiments.
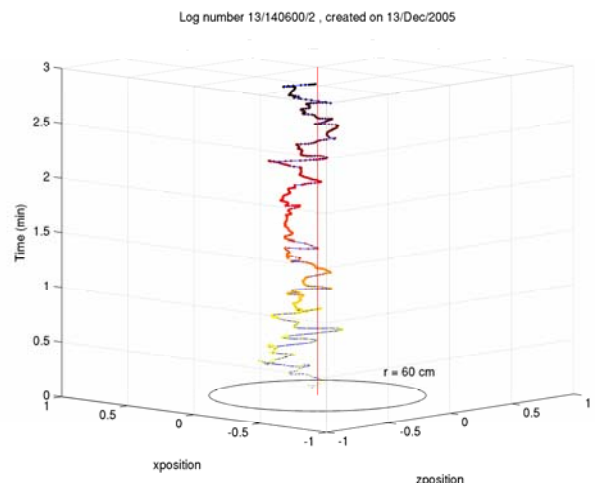

Figure 4: Visualization of the tracker data for one subject exposed to the "visual only" condition.

"Full sequenced" (mean=25.19, st.d=5.91).

The other conditions show a significant reduction in motion.

Tracked movement	Full	Music	Full Seq	Visuals only	Visuals w. footsteps	Sound + 3D
Mean	26,47	20,95	25,19	21,41	22,82	21,77
Median	26,54	20,79	24,31	21,61	25,66	21,87
st.d.	5,6	6,38	5,91	6,39	6,89	6,74

Table 2: Motion analysis for the different conditions.

The significance of the results is outlined in Table 3. In this table, each condition was tested among the others.

	Full	Music	Full seq.	Visual only	Visual w. foot	Sound + 3D
Full						
Music	0.003					
Full seq.	0.243	0.018				
Visual Only	0.006	0.41	0.03			
Visual w. foot	0.04	0.197	0.132	0.26		
Sound + 3D	0.011	0.347	0.048	0.431	0.32	

Table 3: Comparison of the motion analysis for the different conditions.

As can be seen from Table 3, there exists a clear connection between the stimuli. First of all it is interesting to notice that the condition of "Music" elicits the lowest amount of movement, even less than the condition "Visual Only". However, the difference between the condition "Visual Only" and "Music" is not significant (p=0.410), which translates into that we cannot state that using sounds not corresponding to the environment (such as music), should diminish the amount of movement. The fact that music shows less movement indicates that it is important which sound is used. The condition "Music" was in fact used as control condition for this very purpose.

Results also show that footsteps sounds alone do not appear to cause a significant enhancement in the motion of the subjects. When comparing the results of the conditions "Visual only" versus "Visuals w. footsteps" (no significant difference) and the conditions "Full" versus "Sound+3D" (significant difference) there is an indication that the sound of footsteps benefits from the addition of environmental sounds.

This result shows that environmental sounds are implicitly necessary in a virtual reality environment and we assume that their inclusion is important to facilitate
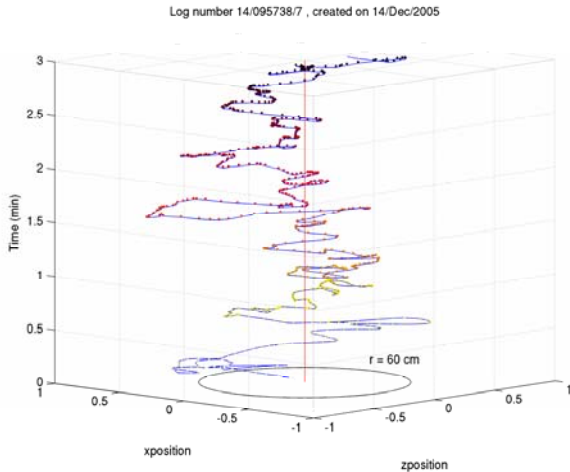

Figure 5: Visualization of the tracker data for one subject exposed to the "full" condition.

Figure 4 shows the visualization of the tracker data over three minutes for one subject exposed to the "Visual only" condition. This visualization was designed to have a way to display in a meaningful way the behavior of the subjects during the experiments.

The circle on the bottom of Figure 4 represents the region of exploration for the visual feedback. Due to the technology used to produce the visual feedback, outside that circle no visual information was present. The vertical line represents the center of the region of exploration. The vertical axis of the figure represents time, while the x and z axis represent the surface where the subjects could walk around. To clarify the visualization, the color information represents the evolution of the motion over time.

We found this visualization rather useful, since it represents clearly quantity of motion of different subjects. Figure 5 shows the same visualization as Figure 4 but in this case for a subject exposed to the condition "Full". Notice how the motion pattern in Figure 4 is significantly less pronounced than in Figure 5.
This is consistent with the test results reported in the following section.

5. Results

Table 2 shows the results obtained by analysing the quantity of motion over time for all subjects for the different conditions. Such analysis was performed by calculating motion over time using the tracker data, where motion was defined as Euclidian distance over time for the motion in 3D.

Our results show that there is clear evidence of the impact of sound in relation to visual only stimuli. As can be seen in Table 2, motion is higher in conditions were the auditory stimuli is "Full" (mean=26.47, st.d=5.6) or

the subjects in accepting the faithfulness of the simulation.

6. Conclusions

In this paper we investigated the role of dynamic sounds in enhancing motion in virtual reality.

Results show that 3D sound with moving sound sources and auditory rendering of ego-motion enhance the quantity of motion of subjects visiting the VR environment.

It very interesting to notice that it is not the individual auditory stimulus that affects the increase of motion of the subjects, but rather that it is the combination of soundscapes, 3-dimensional sound and auditory rendering of one's own motion that induces a higher degree of motion.

We are currently extending these results to environments were the visual feedback is more dynamic and interactive, such as computer games and virtual environments reproduced using 3D graphics.

7. References

[1] Nordahl, Rolf, "Design and evaluation of a multi-modal footstep controller with VR-applications", Proceeding of 2nd International Conference on Enactive Interfaces, Genoa, Italy, 2005.

[2] Cook, P.R. Real sound synthesis for interactive applications. AK Peters, 2002.

[3] Shilling, Russel and Krebs, Eric, "Videogame and Entertainment Industry Standard Sound Design Techniques and Architectures for Use in Videogames, Virtual Environments and Training Systems", MOVES Institute, Naval Postgraduate School.

[4] Freeman, J. and Lessiter, J., "Here, There and Everywhere: The Effects of Multichannel Audio on Presence", Proceedings of the 2001 International Conference on Auditory Display, 2001.

[5] Storms, Russel L. and Zyda, Michael J., "Interactions in Perceived Quality of Auditory-Visual Displays", Presence, Vol. 9, No. 6., December 2000.

[6] Shinn-Cunningham, B., "Applications of Virtual Auditory Displays", Proceedings of the 20th international Conference of the IEEE Engineering in Biology and Medicine Society, Vol. 20, No. 3, 1998.

[7] Larson, Pontus et al., "The Actor-Observer Effect in Virtual Reality Presentations", Cyberpsychology and behaviour, Volume 4, Number 2, 2001.

[8] Kahn, James K. et al., " Integrating Sounds and Motions in Virtual Environments", Presence, v. 7, No.. 1, pp. 67-77, February 1998.

[9] Pulkki,V., "Spatial sound generation and perception by amplitude panning techniques". PhD thesis, Helsinki University of Technology, Espoo, Finland, 2001.

[10] Dimitrov, Smilen, "A simple practical approach to wireless distributed data acquisition", 9th semester project on the Medialogy education, Aalborg University Copenhagen, 2005 .

[11] Mozart, Wolfgang Amadeus, "Piano Quintet in E flat, K. 452, 1. Largo Allegro Moderato", Philips "Digitals Classics, 446 236-2, 1987.

[12] Bordwell, David and Thompson, Kristen., "Film Art", McGraw-Hill Companies; 7th edition (June 1, 2003)

[13] Bakstein, Hynek and Pajdla, Tomas, "Visual Fidelity of Image Based Rendering", Proceedings of the Computer Vision Winter Workshop 2004, 2004

[14] Luciani, A., "Dynamics as a common criterion to enhance the sense of presence in virtual environments", Proceedings of 7th International Conference on Presence, Presence 2004, 96-103. Valencia, Spain, 2004