



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Evaluating Environmental Sounds from a Presence Perspective for Virtual Reality Applications

Nordahl, Rolf

Published in:
Eurasip Journal on Audio, Speech, and Music Processing

DOI (link to publication from Publisher):
[10.1155/2010/426937](https://doi.org/10.1155/2010/426937)

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Nordahl, R. (2010). Evaluating Environmental Sounds from a Presence Perspective for Virtual Reality Applications. *Eurasip Journal on Audio, Speech, and Music Processing*, 2010.
<https://doi.org/10.1155/2010/426937>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Research Article

Evaluating Environmental Sounds from a Presence Perspective for Virtual Reality Applications

Rolf Nordahl

Medialogy, Aalborg University Copenhagen, Lautrupvang 15, 2750 Ballerup, Denmark

Correspondence should be addressed to Rolf Nordahl, rn@media.aau.dk

Received 15 January 2010; Accepted 31 August 2010

Academic Editor: Andrea Valle

Copyright © 2010 Rolf Nordahl. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a methodology to design and evaluate environmental sounds for virtual environments. We propose to combine physically modeled sound events with recorded soundscapes. Physical models are used to provide feedback to users' actions, while soundscapes reproduce the characteristic soundmarks of an environment. In this particular case, physical models are used to simulate the act of walking in the botanical garden of the city of Prague, while soundscapes are used to reproduce the particular sound of the garden. The auditory feedback designed was combined with a photorealistic reproduction of the same garden. A between-subject experiment was conducted, where 126 subjects participated, involving six different experimental conditions, including both uni- and bimodal stimuli (auditory and visual). The auditory stimuli consisted of several combinations of auditory feedback, including static sound sources as well as self-induced interactive sounds simulated using physical models. Results show that subjects' motion in the environment is significantly enhanced when dynamic sound sources and sound of egomotion are rendered in the environment.

1. Introduction

The simulation of environmental sounds for virtual reality (VR) applications has reached a level of complexity that most of the sonic phenomena which happen in the real world can be reproduced using physical principles or procedural algorithms. However, until now little research has been performed on how such sounds can contribute to enhance sense of presence and immersion when inserted in a multimodal environment. Although sound is one of the fundamental modalities in the human perceptual system, it still contains a large area for exploration for researchers and practitioners of VR [1]. While research has provided different results concerning multimodal interaction among the senses [2], several questions remain in how one can utilize to the highest potential audiovisual phenomena when building interactive VR experiences.

As a matter of fact, following the computational capabilities of evolving technology, VR research has moved from being focused on unimodality (e.g., the visual modality) to new ways to elevate the perceived feeling of being virtually present and to engineer new technologies that may offer a

higher degree of immersion, here understood as presence considered as immersion [3].

Engineers have been interested in the audio-visual interaction from the perspective of optimizing the perception of quality offered by technologies [4, 5]. Furthermore, studies have shown that by utilizing audio, the perceived quality of lower quality visual displays can increase [6]. Likewise, researchers from neuroscience and psychology have been interested in the multimodal perception of the auditory and visual senses [7]. Studies have been addressing issues such as how the senses interact, which influences they have on each other (predominance), and audio-visual phenomena such as the cocktail party effect [8] and the ventriloquism effect [9].

The design of immersive virtual environments is a challenging task, and cross-modal stimulation is an important tool for achieving this goal [10]. However, the visual modality is still dominant in VR technologies. A common approach when designing multimodal systems consists of adding other sensorial stimulations on top of the existing visual rendering. This approach presents several disadvantages and does not always allow to exploit the full potential which can be provided by a higher consideration to auditory feedback.

2. Auditory Presence in Virtual Environments

The term presence has been used in many different contexts, and there is still need for the clarification of this term [11]. Such phenomenon has recently been elevated to a status, where it has been used as a qualitative metric for evaluation of virtual reality systems [12]. Most researchers involved in presence studies agree that presence can be defined as a feeling of “being there” [12, 13]. Presence can also be understood as “perceptual illusion of non-mediation” [12] or “suspension of disbelief” of being located in environments that are not real [13].

In [3], Lombard and Ditton outline different approaches to presence. Presence can be viewed as social richness, realism, transportation, and immersion. Sound has received relatively little attention in presence research, although the importance of auditory cues in enhancing sense of presence has been outlined by several researchers [11, 14, 15]. Most of the research relating to sound and presence has examined the role of sound versus nonsound and the importance of spatial qualities of the auditory feedback.

In [16], some experiments were performed with the aim to characterize the influence of sound quality, sound information, and sound localization on users’ self-ratings of presence. The sounds used in their study were mainly binaurally recorded ecological sounds, that is, footsteps, vehicles, doors, and so forth. It was found that especially two factors had high positive correlation with sensed presence: sound information and sound localization.

The previously described research implies that there are two important considerations when designing sounds for VEs, namely, that sounds should be informative and enable listeners to imagine the original (or intended) scene naturally and the other being that sound sources should be well localizable by listeners.

Another related line of research has been concerned with the design of the sound itself and its relation to presence [17, 18]. Taking the approach of ecological perception, in [17] it is proposed that expectation and discrimination are two possibly presence-related factors: expectation being the extent to which a person expects to hear a specific sound in a particular place and discrimination being the extent to which a sound will help to uniquely identify a particular place. The result from their studies suggested that, when a certain type of expectation was generated by a visual stimulus, sound stimuli meeting this expectation induced a higher sense of presence as compared to when sound stimuli mismatched with expectations were presented along with the visual stimulus. These findings are especially interesting for the design of computationally efficient VEs, since they suggest that only those sounds that people expect to hear in a certain environment need to be rendered.

In previous research, we described a system which provides interactive auditory feedback made of a combination of self-sounds and soundscape design [19]. The goal was to advocate the use of interactive auditory feedback as a means to enhance motion of subjects and sense of presence in a photorealistic virtual environment. We focused both on ambient sounds, defined as sound characteristics of

a specific environment which the user cannot modify, as well as interactive sounds of subjects’ footsteps, which were synthesized in real time and controlled by actions of users in the environment. The idea of rendering subjects’ self-sound while walking on different surfaces is motivated by the fact that walking conveys enactive information which manifests itself predominantly through haptic and auditory cues. In this situation, we consider visual cues as playing an integrating role and to be the context of the experiments. In this paper, we extend our research by providing an in-depth evaluation of the system and its ability to enhance the sense of presence and motion of subjects in a virtual environment. We start by describing the context of this research, that is, the BENOGO project, whose goal was to design photorealistic virtual environments where subjects could feel present. We then describe the multimodal architecture designed and the experiments whose goal was to assess the role of interactive auditory feedback in enhancing motion of subjects in a virtual environment as well as sense of presence.

3. The BENOGO Project

Among the different initiatives to investigate how technology can enhance sense of immersion in virtual environments, the BENOGO project (which stands for “being there without going”) (<http://www.benogo.dk>), completed in 2005, had as its main focus the development of new synthetic image-rendering technologies (commonly referred to as Image-Based Rendering (IBR)) that allowed photorealistic 3D real-time simulations of real environments.

The project aimed at providing a high degree of immersion to subjects for perceptual inspection through artificially created scenarios based on real images. Throughout the project, the involved researchers wished to contribute to a multilevel theory of presence and embodied interaction, defined by three major concepts: immersion, involvement, and fidelity. At the same time, the project aimed at improving the IBR technology on those aspects that were found most significant in enhancing the feeling of presence. The BENOGO project was concerned with the reproduction of real sceneries that might be even taken from surroundings familiar to the subject that uses the technology. The thought behind such approach is that in the future we can offer people to visit sites without people having to physically travel to the place.

The BENOGO project makes extensive use of IBR, that is, the photographic reproduction of real scenes. Such technique is dependent on extensive collections of visual data and therefore makes considerable demand on data processing and storage capabilities. One of the drawbacks of reconstructing images using the IBR technique is the fact that, when the pictures are captured, no motion information can be present in the environment. This implies that the reconstructed scenarios are static over time. Depth perception and direction are varied according to the motion of the user, which is able to investigate the environment at 360° inside the so-called region of exploration (REX). However, no events happen in the environment, which make it rather uninteresting to explore.



FIGURE 1: An image of the prague botanical garden used as visual feedback in the experiments.

An occurring problem of IBR technology for VEs has been that subjects in general showed very little movement of head and body. This is mostly due to the fact that only visual stimuli were provided. By transferring information from film studies and current practice, practitioners emphasize that auditory feedback such as sound of footsteps signifies the characters giving them weight and thereby subjecting the audience to interpretation of embodiment.

We hypothesize that the movement rate can be significantly enhanced by introducing self-induced auditory feedback produced in real time by subjects while walking in the environment.

We start by describing the content of the multimodal simulation, and we then describe how the environment was evaluated.

4. Designing Environmental Sounds for Virtual Environments

The content of the proposed simulation was a reproduction of the Prague botanical garden, whose visual content is shown in Figure 1. As seen in Figure 1, the environment has a floor made of concrete, where subjects are allowed to walk. This is an important observation when sonically simulating the act of walking in the environment.

The main goal of the auditory feedback was both to reproduce the soundscape of the botanical garden of Prague and to allow subjects to hear the sound of their own footsteps while walking in the environment. The implementation of the two situations is described in the following.

4.1. Simulating the Act of Walking. We are interested in combining sound synthesis based on physical models with soundscape design in order to simulate the act of walking on different surfaces and place them in a context. Specifically, we developed real-time sound synthesis algorithms which simulate the act of walking on different surfaces. Such sounds

were simulated using a synthesis technique called modal synthesis [20].

Every vibrating object can be considered as an exciter which interacts with a resonator. In our situation, the exciters are the subjects' shoes, and the resonators are the different walking surfaces. In modal synthesis, every mode (i.e., every resonance) of a complex object is identified and simulated using a resonator. The different resonances of the object are connected in parallel and excited by different contact models, which depend on the interaction between the shoes and the surfaces. Modal synthesis has been implemented to simulate the impact of a shoe with a hard surface.

In the case of stochastic surfaces, such as the impact of a shoe with gravel, we implemented the physically informed stochastic models (PhISM) [21].

The footstep synthesizer was built starting by analyzing footsteps recorded on surfaces obtained from the Hollywood Edge Sound Effects library (<http://www.hollywoodedge.com>). For each recorded set of sounds, single steps were isolated and analyzed. The main goal of the analysis was to identify an average amplitude envelope for the different footsteps, as well as extracting the main resonances and isolating the excitation.

A real-time footstep synthesizer, controlled by the subjects using a set of sandals embedded with force sensors was designed. Such sandals are shown in Figure 2. By navigating in the environment, the user controlled the synthetic footsteps sounds.

Despite its simplicity, the shoe controller was effective in enhancing the user's experience, as it will be described later. While subjects were navigating around the environment, the sandals were coming in contact with the floor, thereby activating the pressure sensors. Through the use of a microprocessor, the corresponding pressure value was converted into an input parameter which was read by the real-time sound synthesizer implemented in Max/MSP (<http://www.cycling74.com>). The sensors were wirelessly connected to a microcontroller, as shown in Figure 2, and the microprocessor was connected to a laptop PC.

The continuous pressure value was used to control the force of the impact of each foot on the floor, to vary the temporal evolution of the synthesized generated sounds. The use of physically based synthesized sounds allowed to enhance the level of realism and variety compared to sampled sounds, since the produced sounds of the footsteps depended on the impact force of subjects in the environment, and therefore varied dynamically. In the simulation of the botanical garden, we used two different surfaces: concrete and gravel. The concrete surface was used most of the time and corresponded to the act of walking around the visitors' floor. The gravel surface was used when subjects were stepping outside the visitors' floor.

Both surfaces were rendered through an 8-channel surround sound system.

4.2. Simulating Soundscapes. In order to reproduce the characteristic soundmarks of a botanical garden, a dynamic soundscape was built. The soundscape was designed by

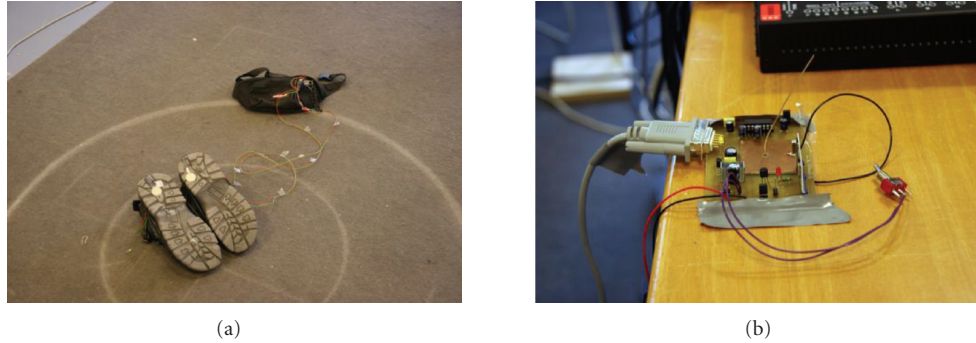


FIGURE 2: The sandals (a) enhanced with pressure-sensitive sensors wirelessly connected to a microprocessor (b).

creating an 8-channel soundtrack in which subjects could control the position of different sound sources.

In the laboratory shown in Figure 4, 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 method, 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 [22], where eight-channel surround is presently implemented, in order to perform in the future experiments with higher-quality visual feedback. This is the reason why 8-channel sound rendering was chosen compared to, for example, binaural rendering [23].

Three kinds of auditory feedback were implemented:

- (1) “static” soundscape, reproduced at max. peak of 58 dB, measured c-weighted with slow response. This soundscape was delivered through the 8-channel system;
- (2) dynamic soundscape with moving sound sources, developed using the VBAP algorithm, reproduced at max. peak of 58 dB, and 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 [24]).

The content of the soundscape in the first two conditions was the same. The soundscape contained typical environmental sounds present in a garden such as bird singing and insects flying. The soundscape was designed by performing a recording in the real botanical garden in Prague and reproducing a similar content by using sound effects from the Hollywood Edge Sound Effects library.

In the first and second conditions, the soundscape only varied in the way it was rendered. In the second condition, in fact, the position of the sound sources was dynamic and controlled by the user’s motion, who was wearing a head tracker as described below. In the third condition, the dynamic



FIGURE 3: A subject navigating in the virtual environment wearing a head-mounted display (HMD).

soundscape was augmented with auditory simulation of ego-motion obtained by having subjects generating in real-time footsteps of themselves walking in the garden.

5. A Multimodal Architecture

In order to combine the auditory and the visual feedback, together with the shoe controller, two computers were installed in the laboratory. One computer was running the visual feedback and other one the auditory feedback together with the interactive shoes. A Polhemus tracker (IsoTrak II3), 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 via TCP socket. 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, allowed detecting the footsteps of the subjects and mapping these to the real-time sound synthesis engine. The different hardware components were connected together as shown in Figure 6.

The visual stimulus was provided by a standard PC running SUSE Linux 10. This computer was running the

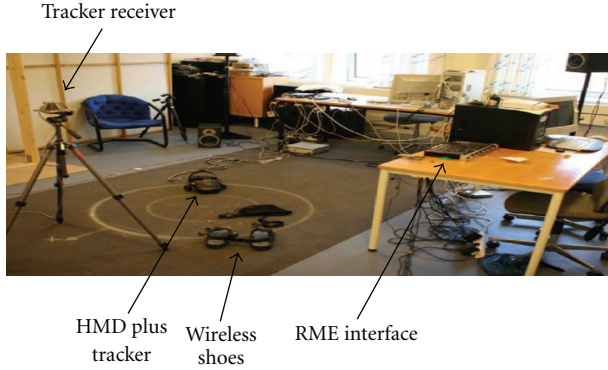


FIGURE 4: A view of the lab setup, where the experiments were run. Notice the two computers, placement of speakers (top/bottom), the HMD (lying on the floor), the tracking receiver (outside the REX), and the sandals.



FIGURE 5: A different view of the 8-channels surround sound lab, where the experiments were run.

BENOGO software using the REX disc called *Prague Botanical Garden*.

The head-mounted-display (HMD) used was a VRLogic V82. It features Dual 1.3 diagonal Active Matrix Liquid Crystal Displays with resolution per eye: $((640 \times 3) \times 480)$, (921,600 color elements) equivalent to 307,200 triads. Furthermore, the HMD provides a field of view of 60° diagonal. The tracker used (Polhemus IsoTrak II3) provides a latency of 20 milliseconds with a refresh rate of 60 Hz.

The audio system was created using a standard PC running MS Windows XP SP 2. All sound was run through Max/MSP, and as output module a Fireface 800 from RME5 (<http://www.rmeaudio.com/english/firewire/>) was used. Sound was delivered by eight Dynaudio BM5A speakers (<http://www.dynaudioacoustics.com>). Figure 5 shows a view of the surround sound lab, where the experiments were run. In the center of the picture, the tracker's receiver is shown.

6. Evaluating the Architecture

In order to assess how the different kinds of auditory feedback affected users' behavior in the environment, an

experiment was run, where 126 subjects took part. All subjects reported normal hearing and visual conditions. Figure 3 shows one of the subjects participating in the experiment. Before entering the room, subjects were asked to wear a head-mounted display and the pair of sandals enhanced with pressure-sensitive sensors. Subjects were not informed about the purpose of the sensor-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.

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

- (1) Visual only: This condition had only unimodal (visual) input.
- (2) Visual with footstep sounds: In this condition, the subjects had bi-modal perceptual input (audio and visual) comparable to our earlier research [24].
- (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 and 3D sound (using the VBAP algorithm) as well as rendering sounds from ego-motion (the subjects triggered sounds via their footsteps).
- (4) Visual with fully 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 and 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.
- (5) Visual with sound +3D sound: This condition introduced bi-modal (audio and 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 music described before (see [25]). 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 deduce how the specific variable sound design from the other experimental conditions affects the subjects.

Subjects were randomly assigned to one of the six conditions above. The six different conditions, together with information about the subjects, are summarized in Table 1.

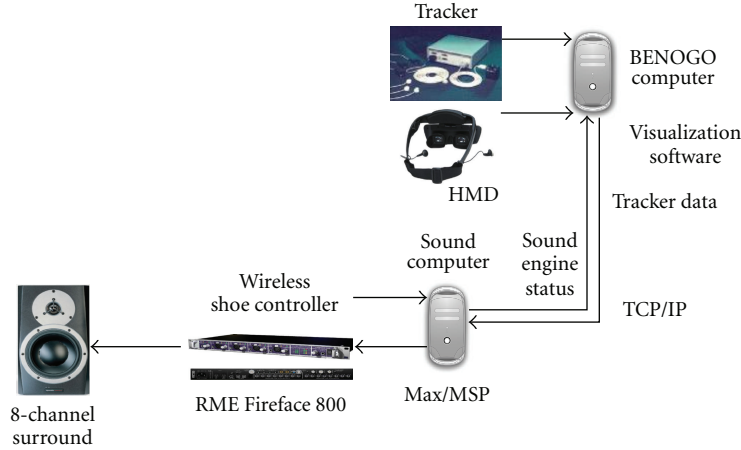


FIGURE 6: Connection of the different hardware components in the experimental setup.

TABLE 1: Six different conditions to which subjects were exposed during the experiments. The number in the second column refers to the auditory feedback previously described.

Condition	Auditory stimuli	No. subj.	Mean (age)	St. d. (age)
Visual only	None	21	25.6	4.13
Visual 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		21	28	8.1

TABLE 2: Motion analysis for the different conditions considering only the 2D motion.

Tracked movement	Mean	Median	St. d.
Visual only	21.41	21.61	6.39
Visual w. foot	22.82	25.66	6.89
Full	26.47	26.54	5.6
Full Seq.	25.19	24.31	5.91
Sound + 3D	21.77	21.87	6.74
Music	20.95	20.79	6.39

7. 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 Euclidean distance from the starting point position over time for the motion in 2D. Since motion was

derived from the tracker's data placed on top of the head mounted display, only the motion of the head of the subjects was tracked. In particular, Table 2 shows data obtained by analyzing the motion of the subjects in the horizontal plane. It is interesting to notice how the condition *Music* elicits the lowest amount of movement (mean = 20.95), even less than the condition *Visual only* (mean = 21.41).

The significance of the results is outlined in Table 3, where the corrected P -value was calculated for the different conditions, using a t -test. The difference between the conditions *Visual only* and *Music* is not significant ($P = .410$), 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 the content of the sound used is important. 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 *Visual w. foot* (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 motion. This is an important observation which is validated in the real world, when we are used to perceive our self-sound always in the context of the surrounding space.

We additionally analyzed the motion of the subjects taking into account also the vertical movement, which represents the action of subjects standing or going down on their knees. Such action was performed by several subjects when trying to locate objects in the lower part of the environment. Results are shown in Table 4.

As Table 4 shows, results are very consistent with the analysis and results without taking into account the vertical motion. The trends, seen from the condition ranked according to mean values, indicate that the addition of auditory

TABLE 3: Comparison of the 2D motion analysis for the different conditions (P -value).

	Visual only	Visual w. foot	Full	Full seq.	Sound + 3D	Music
Visual only						
Visual w. foot	0.26					
Full	0.006	0.04				
Full seq.	0.03	0.132	0.243			
Sound + 3D	0.431	0.32	0.022	0.048		
Music	.41	.197	.003	.018	.347	

TABLE 4: Motion analysis for the different conditions including vertical movement.

Tracked movement	Mean	Median	St.d.
Visual only	33.23	33.51	9.71
Visual w. foot	35.65	38.13	8.63
Full	40.93	41.05	7.9
Full seq.	38.14	37.08	8.82
Sound + 3D	33.59	33.96	10.27
Music	31.92	30.81	9.38

TABLE 5: Average presence index for the six experimental conditions.

Presence index	Mean	Median	St.d.
Visual only	4.58	4.5	0.92
Visual w. foot	4.82	5	1.06
Full	4.77	4.75	1.08
Full Seq	4.79	4.75	0.69
Sound + 3D	4.81	5	0.79
Music	4.82	5	1.13

stimuli induces a positive effect on motion. Both for head and complete movement, results show that the mean values for the conditions are similar in ranking. A statistical analysis shows that in the conditions *Full* and *Full seq.* when viewed against the condition *Visual only*, the average body motion is significantly higher when the auditory stimuli are introduced. (*Full* compared to *Visual only* ($P = .005$), *Full seq* compared to *Visual only* ($P = .051$)).

Figures 7 and 8 show the Polhemus tracker data over time for one subject in the 2D plane with the six different conditions, with three conditions represented for each figure.

The circle at the bottom of the tracker data represents the REX. The fact that subjects are allowed to move freely in the space prevents us from visualizing the path of each subject, or an average of the different paths. However, we chose some characteristic behavior of the different conditions, and we also noticed that a similar behavior can be seen also in subjects in the same condition. The most striking feature in the plots is the fact that the limited amount of motion in the condition with only visual feedback (Figure 7(a)) is clearly noticeable. The subject in the full condition (Figure 7(c)) appears to be interested in an active exploration of the environment. The same can be said for the subject in the condition visual plus footsteps (Figure 7(b)).

8. Measuring Presence

As a final analysis of the six experimental conditions, we investigated the qualitative measurements of the feeling of presence. Through the tests for all conditions we implemented all questions from the SVUP questionnaire [26]. The SVUP is concerned with examining four items, where the most important item in relation to our thesis is the feeling of presence. The SVUP questionnaire does so by asking the subjects to answer four questions which all relate to the feeling of presence. The results of these answers are then averaged for each subject, resulting in what is referred to as the presence index. The questions relate to the naturalness of interaction with the environment and sense of presence and involvement in the experience. All answers were given on a Likert scale [27], from 1–7, (1 represents not at all and 7 represents very much).

Table 5 shows the results of the presence questionnaire for the different conditions. The first thing to notice is that all the conditions with auditory feedback have a higher presence rate than the condition with only visuals. This result confirms previous research which showed that auditory feedback enhances sense of presence.

It is also interesting to notice the answers to one of the questions from the SVUP questionnaire, namely, how much subjects felt that the experience was influenced by their own motion, rated on a scale from 1 to 100. The condition visuals w. footsteps has the highest rating in this situation (mean = 83.05), with a significant difference with the second

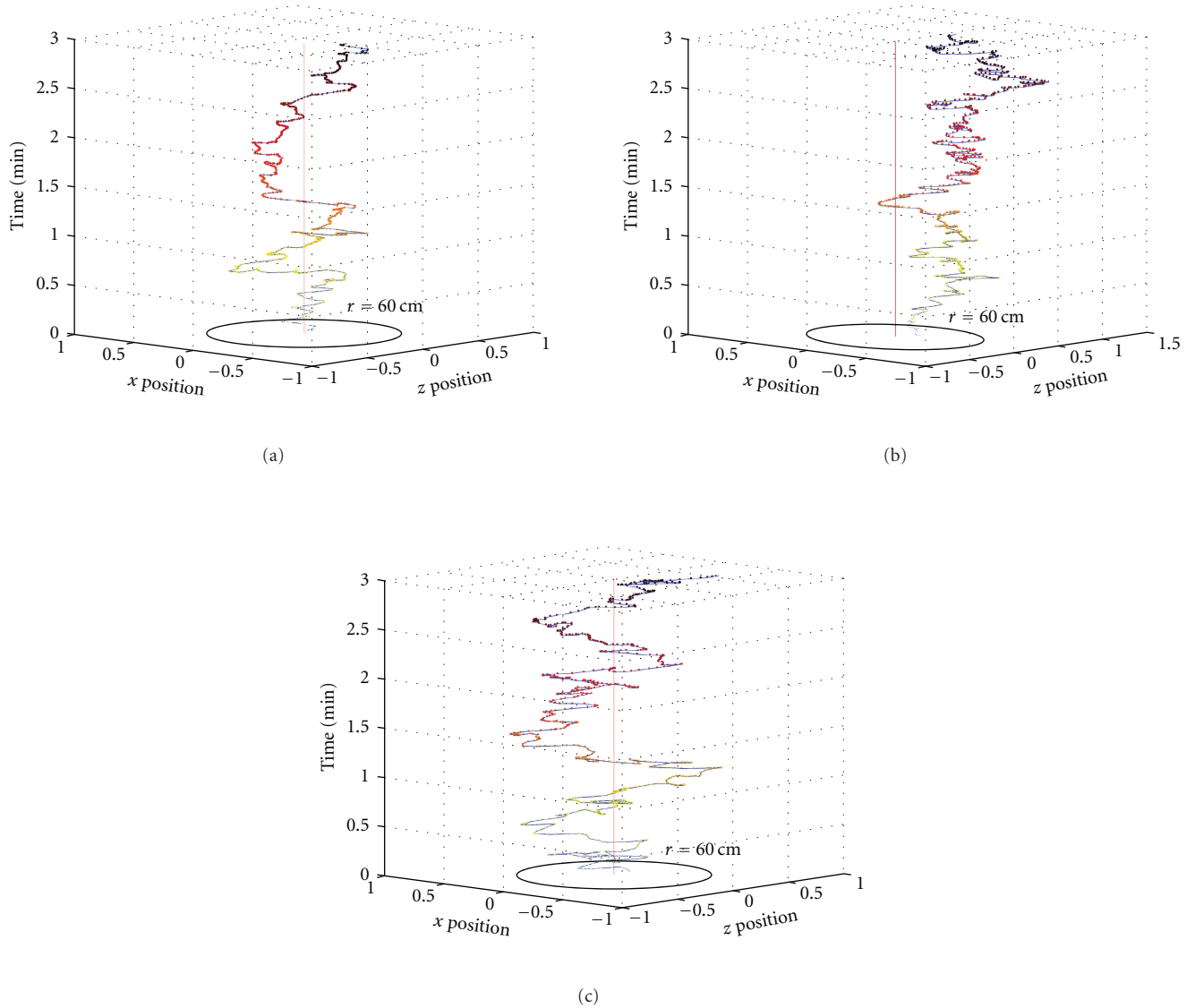


FIGURE 7: Visualization over time of the motion of one subject in the six different conditions. From (a) to (c): visual, visual w. foot, and full.

highest ranked condition in the list (full seq., mean = 71.4) ($P < .02$). This shows that the footstep synthesizer actually works, since users realize that they are controlling the feedback. Moreover, it is reasonable to assume that, when no soundscape is present, the users can focus more attention on the footstep sounds, therefore, recognizing the tight coupling between the act of walking and footsteps sounds in the environment.

An overall analysis of variance on the results shows that no significant differences were noticeable among the different conditions.

One reason that may affect the overall results derived from the self-report of the subjects is that the experiments of this study were done as a between subjects exploratory study. The fact that the individual subject only experienced one condition was optimal in the sense that issues concerning

subjects becoming accustomed to the VE or finding it increasingly boring was minimized.

However, since the subjects have no other conditions as a frame of reference, this may be a plausible cause of what we have experienced through these results of the SVUP presence index, that is, that between-subjects as a method for this particular presence index is not adequate since the subjects give their initial feeling of how they felt without having anything to measure this feeling against. However, the quantitative data from the motion tracking shows clear results with significance, and the between-subjects strategy is well suited towards such experiments. Overall, mean and median values are very central in the scale, with a small standard deviation, which means that users provided in general an average evaluation, without any specific condition which was significantly more pronounced in the Likert scale.

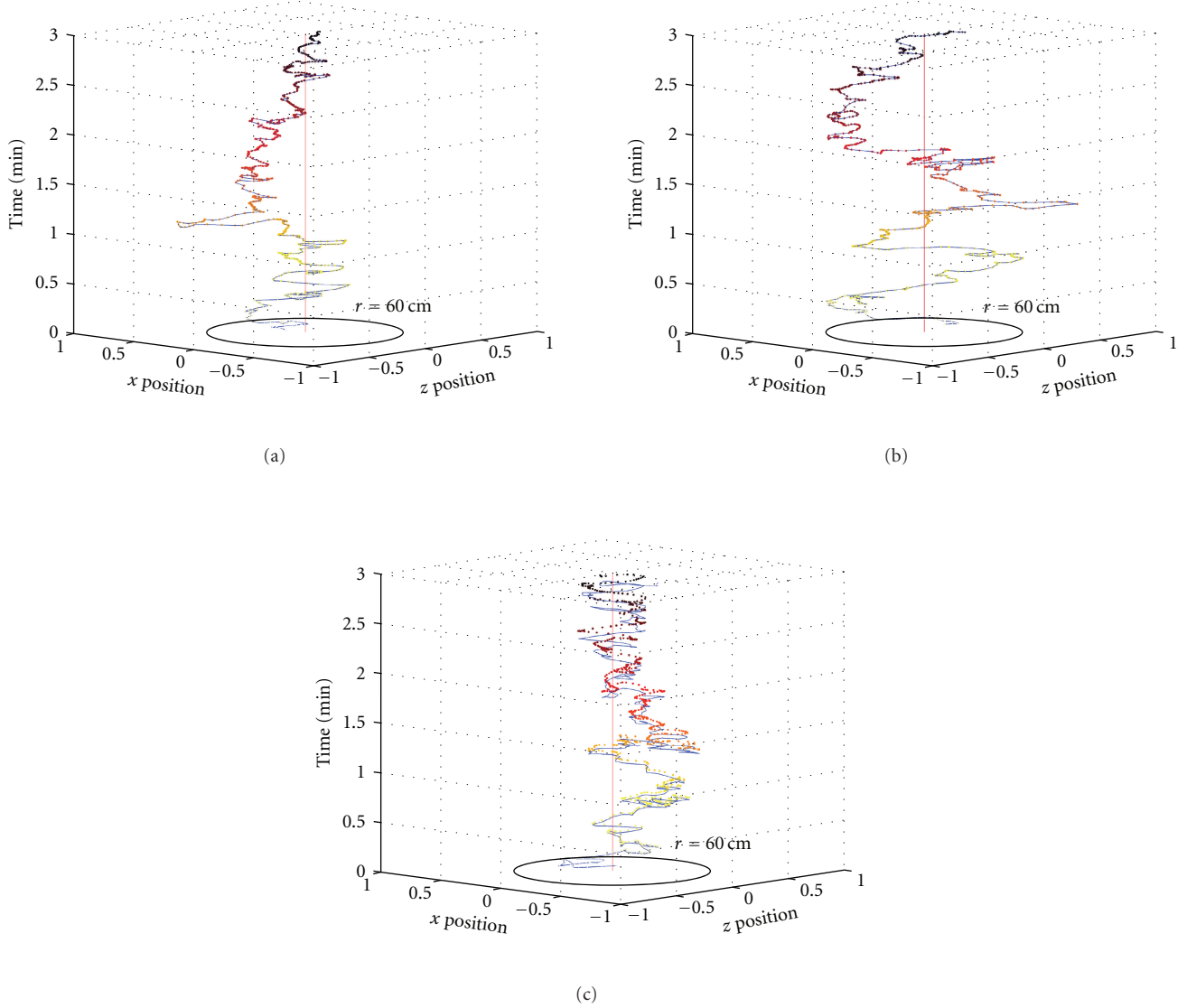


FIGURE 8: Visualization over time of the motion of one subject in the six different conditions. From (a) to (c): full sequenced, sound + 3D, and music.

This can be due to the fact that subjects experienced only one condition, so they did not have a frame of reference to compare.

9. Conclusion

In this paper, we investigated the role of dynamic sounds in enhancing motion and presence in virtual reality. Results show that 3D sounds with moving sound sources and auditory rendering of ego-motion significantly enhance the quantity of motion of subjects visiting the VR environment.

It is very interesting to notice that it is not the individual auditory stimulus that affects the increase of motion of the subjects, but rather 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 also investigated whether the sense of presence was increased when interactive sonic feedback was provided to the users. Results from the SVUP presence questionnaire do not show any statistical significance in the increase of presence.

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

Disclosure

Permission to make digital or hard copies of all or part of this paper for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice

and the full citation on the first page. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

References

- [1] K. M. Stanney et al., *Handbook of Virtual Environments: Design, Implementation, and Applications*, Lawrence Erlbaum Associates, 2002.
- [2] B. E. Stein, M. A. Meredith, and S. Wolf, *The Merging of the Senses*, The MIT Press, Cambridge, Mass, USA, 1993.
- [3] M. Lombard and T. Ditton, "At the heart of it all: the concept of presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, 1997.
- [4] N. F. Dixon and L. Spitz, "The detection of auditory visual desynchrony," *Perception*, vol. 9, no. 6, pp. 719–721, 1980.
- [5] S. Rihs, "The influence of audio on perceive picture quality and subjective audio-video delay tolerance," in *Proceeding of the MOSAIC Workshop Advanced Methods for the Evaluation of Television Picture Quality*, 1995.
- [6] R. L. Storms and M. J. Zyda, "Interactions in perceived quality of auditory-visual displays," *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 6, pp. 557–580, 2000.
- [7] A. Kohlrausch and S. Vand de Par, "Auditory-visual interaction: from fundamental research in cognitive psychology to (possible) applications," in *Proceedings of the IST/SPIE Conference on Human Vision and Electronic Imaging*, 1999.
- [8] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice*, vol. 12, 1992.
- [9] S. Handel, *Perceptual Coherence: Hearing and Seeing*, Oxford University Press, Oxford, UK, 2006.
- [10] N. Durlach and A. Mavor, *Virtual reality: Scientific and Technological challenges*, National Academy Press, Washington, DC, USA, 1995.
- [11] M. Slater, "A note on presence terminology," *Presence Connect*, vol. 3, no. 3, 2003.
- [12] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: the ITC-sense of presence inventory," *Presence*, vol. 10, no. 3, pp. 282–297, 2001.
- [13] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments. A presence questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, 1998.
- [14] W. A. IJsselstein, *Presence in Depth*, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2004.
- [15] R. H. Gilkey and J. M. Weisenberger, "The sense of presence for the suddenly deafened adult—implications for virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 4, pp. 357–363, 1995.
- [16] K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone, "Psychological factors involved in auditory presence," *Acoustical Science and Technology*, vol. 24, no. 1, pp. 42–44, 2003.
- [17] P. Chueng and P. Marsden, "Designing auditory spaces to support sense of place: the role of expectation," in *Proceedings of the CSCW Workshop: The Role of Place in Shaping Virtual Community*, Citeseer, 2002.
- [18] S. Serafin and G. Serafin, "Sound design to enhance presence in photorealistic virtual reality," in *Proceedings of the International Conference on Auditory Display*, pp. 6–9, Citeseer, 2004.
- [19] R. Nordahl, "Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion," *Proceedings of Presence*, pp. 57–62, 2006.
- [20] J. M. Adrien, "The missing link: modal synthesis," *Representations of Musical Signals Table of Contents*, pp. 269–298, 1991.
- [21] P. R. Cook, "Physically informed sonic modeling (PhISM): synthesis of percussive sounds," *Computer Music Journal*, vol. 21, no. 3, pp. 38–49, 1997.
- [22] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, "Cave. Audio visual experience automatic virtual environment," *Communications of the ACM*, vol. 35, no. 6, pp. 65–72, 1992.
- [23] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*, AP Professional, Boston, Mass, USA, 1994.
- [24] R. Nordahl, "Auditory rendering of self-induced motion in virtual reality," M. Sc. project report, Department of Medialogy, Aalborg University Copenhagen, 2005.
- [25] Mozart and W. Amadeus, "Piano Quintet in E flat, K. 452, 1. Largo Allegro Moderato," Philips Digitals Classics, 446 236-2, 1987.
- [26] D. Västfjäll, P. Larsson, and M. Kleiner, "Development and validation of the Swedish viewer-user presence questionnaire (SVUP)," 2000.
- [27] T. J. Maurer and H. R. Pierce, "A comparison of likert scale and traditional measures of self-efficacy," *Journal of Applied Psychology*, vol. 83, no. 2, pp. 324–329, 1998.