AALBORG UNIVERSITY

DENMARK

Presence studies as an evaluation method for user experiences in multimodal virtual environments

Nordahl, Rolf

Publication date:
2010

Document Version
Tidlig version også kaldet pre-print

PH.D. DISSERTATION

Presence studies as an evaluation method for user experiences in multimodal virtual environments

ROLF NORDAHL

Department of Architecture, Design and Media Technology
AALBORG UNIVERSITY COPENHAGEN
Denmark

Presence studies as an evaluation method for user experiences in multimodal virtual environments

Rolf Nordahl

Department of Architecture, Design and Media Technology
Medialogy

Aalborg University Copenhagen
2750 Ballerup
Phone: +45 99402472

Author e-mail: `rn@media.aau.dk`

Pursuing Rainbows

For Emma Ilena Julia Nordahl

Dear Emma,

In lack of words to describe the amazement of life, existence, being and creation of the world and life I would like to use parts of a personal letter that Jacques-Yves Cousteau wrote to his son, Philippe Cousteau Sr. - which later became the vision and dream to pursue for his grandson.

I hope you some day will find the same inspiration of the wonders of our world and perhaps give me the joyful gift to be my guide to the secrets of your visions, hopes and dreams.

"I will always remember that day of July 1963 when you joined the "Conshelf II" expedition along the Shab Rumi reef, in the Red Sea. The sun was setting when you climbed onboard the Calypso from the launch that had driven you from Port Sudan airport. But I would not give you time to relax I was too impatient to show you our "village under the sea" before it became too dark. Hastily, we both donned our aqualungs, and slowly, sensually, we submerged into the welcoming water, as warm as our blood. When we started for an unforgettable stroll with slow strokes of our long stretched legs and breathing deep lungfuls of air I kept your hand in mine to guide you from "Starfish house" where six oceanauts were having dinner, to the onion shaped "diving saucer garage", to the "tool house", the "fish farm", to the "deep cabin" where we observed the two "black masked" oceanauts go to bed...and to the anti shark cages strewn here

and there as emergency shelters. I introduced you to Jules, the great barracuda who had adopted us. I showed you the cave in which the large "bump-fish" went to sleep at night, and of course, we met the inevitable sharks who kept cruising around the village. Twilight was turning to sheer darkness, our structures became eerie shadows, the fish were just moving pieces of the sea. I was still holding your hand when we returned to the ladder; I felt strangely proud not of what we had achieved, but because our dreams were always shared so intimately. Three years ago, I found myself sitting near you in the cockpit of our Catalina, the seaplane you had equipped especially for oceanography and for diving. From years of gliding, handgliding, piloting planes and helicopters, ballooning, you had acquired an unusual expertise. Now you were giving me a ride to the Mexican island of Isabella, in the Pacific. Taking off in sheaves of water, the whole of the plane was an extension of your body, the roar of the motors was the expression of your joy, the clouds that dotted your sky were just other forms of water like our own flesh. I look at you, my guide in the sky as I have been your guide in the sea.

I saw your shining face, proud to have something to give back to me, and I smiled, because I knew that pursuing rainbows in your plane, you would always seek after the vanishing shapes of a better world."

I love you, Rolf

Rolf Nordahl
Copenhagen, Denmark, August 2010

Acknowledgements

In gratitude

▷ Erik Granum

▷ Lise Busk Kofoed

▷ Flemming K. Fink

▷ Stefania Serafin

▷ Emma Ilena Julia Nordahl

▷ My mother and my aunt - in loving memory; may you both have found the peace that life could not offer

▷ Tubas Løkkegaard

▷ Dannie Korsgaard

▷ Børge, Ingrid og Maja Frederiksen

<div align="right">
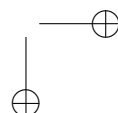
Rolf Nordahl

Copenhagen, Denmark, August 2010

</div>

iv   ACKNOWLEDGEMENTS

Contents

Part I

INTRODUCTION

1

Introduction

In the last decades, research on multisensory interaction has seen a growing attention. However, the applications of findings in this field to Virtual Reality (VR), virtual environments (VE) and multimedia is still an ongoing process. Most of the research has focused on rendering different sensory inputs optimized separately. Moreover, most of the research efforts have been devoted to the visual modality, or the interaction between vision and audition or vision and touch.

Evaluating how users experience complex virtual environments has shown to be a complex task. This is especially true when there is a wish to understand to what degree users are able to feel present or immersed in the environment. Throughout the Ph.D. education in Medialogy it has been my personal ambition to gain insight into if and how presence research can be adapted to the evaluation of different innovative multimodal

environments. By innovative, it is meant environments which are different from the ones which are usually adopted in presence research, such as virtual environments which use commercial software and interfaces.

Being a student at Aalborg University gave me the opportunity to develop projects alongside with state-of-the-art research environments. One such research project that ended during the beginning of this Ph.D. dissertation was the BENOGO project,[1] headed by the Computer Vision and Media Technology (CVMT) group at Aalborg University. Research within that particular project was concerned with using Image Based Rendering techniques (IBR) for rendering static images containing depth information [1]. As it will be explained in details later, while the researchers in the BENOGO project were extremely skilled in visualization and computer graphics, they lacked some sound expertise, a topic where I could contribute.

When the BENOGO project ended, I pursued my research on auditory rendering for virtual environments. At the same time, I applied results from presence research to other fields such as computer games and movies, developing novel evaluation methods based on presence. The investigations into presence research have been tightly connected to my pedagogical activities. Indeed, I have been deeply involved in establishing the Medialogy education in Copenhagen, and among other things developing a connection between research and education.

In October 2008 another research project started, named the Natural Interactive Walking (NIW) project.[2] This project, in some aspects, can be seen as a continuation of the BENOGO project as it pursues some of the same goals in exploring how people can perceive realistic simulations of virtual environments. As stated in the official description, the goal of the NIW project is to investigate possibilities for the integrated and interchangeable use of the haptic and auditory modality in floor interfaces, and for the synergy of perception and action in capturing and guiding human walking. Its objective is to provide closed-loop interaction paradigms, negotiated with users and

---

[1] www.benogo.dk
[2] www.niwproject.eu

validated through experiments, enabling the transfer of skills that have been previously learned in everyday tasks associated to walking, and where multi-sensory feedback and sensory substitution can be exploited to create unitary multimodal percepts.

As it will be described in details later, while one of the major constraints of the BENOGO project is the fact that subjects are exposed to a static rendering, NIW encourages foot-based interactions by rendering actions with visual, auditory and haptic feedback.

This dissertation also benefits from several interactions with students enrolled in the Medialogy education. As coordinator of studies for Medialogy, and while establishing the Medialogy education, I have been interested in finding a red thread among the different semesters. I soon noticed that, no matter which innovative technologies students were involved with, a common occurring issue was how to evaluate them, and how to assess if they contributed in creating a sense of engagement and immersion of the user. I found the field of presence research as an ideal red thread for evaluating multimodal systems, and participated and presented regularly to the International Society of Presence Research (ISPR) annual conference since then. One of the advantages of this field is that researchers come from different backgrounds, some more focused on developing technologies which enhance sense of presence, others learning how to evaluate such technologies. Presence research has been traditionally connected to virtual reality and virtual environments. Traditionally, research on virtual reality has utilized rather expensive commercial equipment mostly for tracking and visualization purposes, such as head mounted displays and 3-dimensional position trackers.

However, in recent years, advances in sensors technology and user interface design has allowed a wide development of low-cost custom made interfaces and input devices which are used to control auditory and visual effects. As an example, the field of sonic interaction design has started to emerge. Sonic Interaction Design is an interdisciplinary area of research and practice that explores ways in which sound can be used to convey information, meaning, and aesthetic and emotional qualities in interactive contexts [2].

1.1   Objectives and questions for the thesis

In this dissertation, novel evaluation techniques for media technology applications based on presence research are proposed. One of the motivations is to explore the possibilities of adopting presence methods and evaluation techniques in fields where they have been unexplored, such as the field of sonic interaction design and the field of computer games.

The main goal is to create new methods for evaluation of user experience in multi-modal environments. Thereby the method becomes the research object of interest.

Concerning the use of interactive sound in virtual environments, it has not yet been a main focus in virtual environments research. Indeed, 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 [3, 4]. 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. Moreover, research in interactive sound driven by alternative interfaces has mostly focused on their engineering side, but little focus has been placed on their evaluation. The same can be said about computer games, where research has also focused mostly on the design and implementation, rather than on evaluation techniques.

The work presented in this thesis is focused on exploring evaluation of sonic interaction through presence techniques as well as proposing the use of presence as an evaluation tool for game development. To encourage this exploration, I custom made interfaces to support and enhance my presence research. Although I have also worked on different applications of novel interfaces and devices, my focus has always been towards their evaluation - and not the development of the technology itself.

General goals of this dissertation are the following:

1. To gain increased knowledge of how presence research can be utilized in the field of sonic interaction design. I have designed multimodal environments which implement interactive sonic feedback, to investigate which role sound plays in

enhancing the sense of presence. Specifically, an aim was to investigate whether interactive self-sound produced by subjects when walking in a virtual environment enhances sense of presence.

2. To innovate in the field of computer games evaluation, I proposed several evaluation techniques based on presence research. Such techniques utilise presence in ways which has never been done before.

3. To extend traditional presence research, which uses commercially made interfaces, I investigated whether the use of custom made devices can contribute to provide new insights into evaluation techniques. To achieve this goal, I have designed several custom made interfaces which will be described in the following chapters.

4. To investigate the validity of existing presence questionnaires, which are a largely discussed topic in the field of presence research, I researched how to adapt and combine several existing questionnaires, as well as qualitative and quantitative test-methodologies. This has been done to explore whether it would contribute to an increased knowledge and outcome of the experiments.

While the field of evaluation techniques and presence research is already multi-faceted and approached by academics from a variety of different traditions, it has been my aspiration to explore whether the areas of evaluation techniques could benefit from adopting techniques from fields slightly estranged to them. In other words, I used existing methodologies in new contexts, aspiring to a new multidisciplinary marriage of evaluation techniques where presence might shed light on issues on the users' experience that could or should be of interest for the academic field.

In this dissertation, several approaches to utilize presence research as a method to evaluate users' experiences in virtual environments are proposed. In particular, presence is investigated in connection with disciplines where it was not applied before, such as sonic interaction design, computer games and movies.

The dissertation is composed of a collection of eight peer-reviewed papers and

an introduction. The papers have been published in International Journals and international conference proceedings. Papers I to V investigate the relationship between auditory rendering and presence. Papers VI to VII adapt presence measuring methods to the evaluation of games and movies.

2

Defining presence

Hand in hand with the computational capabilities of evolving technology, VR research has slowly moved from being focused on uni-modality (e.g., the visual modality). From a status where the visual predominance reigned, the research community has begun to look for new ways to elevate the perceived feeling of being virtually present and to engineer new technologies that may offer a higher degree of immersion. The perceived feeling of virtually being in a simulated environment is usually referred as the feeling of presence.

The nature of presence research is interdisciplinary; ranging from the role of engineers implementing virtual environments, to psychologists and neuroscientists whose goal is to understand how users experiencing such environments will react. In this dissertation, it is my interest to understand to what extent presence research can be

applied to fields which have been until now rather unexplored, such as the interaction with sonic feedback, and the role of presence in games and movies. VR poses both questions and new possibilities, not only for the auditory modality but with multimodal interaction in mind. Engineers have been interested in the audio-visual interaction from the perspective of optimizing the perception of quality offered by technologies [5, 6]. Furthermore, studies have shown that by utilizing audio, the perceived quality of lower quality visual displays can increase [7]. Likewise, researchers from neuroscience and psychology have been interested in the multimodal perception of the auditory and visual senses [6, 8]. Studies have been addressing issues such as how the senses interact, which influences they have on each others, and audio-visual phenomena such as the cocktail party effect [9] and the ventriloquism effect.

Since this dissertation largely deals with measure and evaluation techniques for interactive virtual environments, it is important to understand what the common purposes of such applications are. Besides the obvious qualities that can be developed and capitalized by industry for entertainment purposes, a large part of VR-research is focused on conceptualizing and understanding what defines the perception of presence. Where people in the real world are never confused with being present in an environment, the feeling of *being there* is however difficult when people are situated in mediated environments. From being something regarded as normal and necessary in the real world, the phenomenon of presence has now been elevated to a state where it has been used as qualitative metric for evaluation purposes [10].

While it should pose no discussion that VR deals with mediated experiences, the term presence however introduces unknowns in the definition and understanding of what the word (and the aspects that follow it) covers. A majority of researchers involved in presence agree that presence can be defined as a feeling of *"being there"* [10]. Further analysis of the phenomenon can however lead to variances that may be of importance when trying to understand the purposes and conclusions of various studies in the field. For example, it has been argued that presence should be understood as the *perceptual illusion of non-mediation*, referenced in [10] or *"the suspension of disbelief"* of being located in environments that are not real, referenced in [11].

In [12], Lombard outlines the different approaches to presence. Such approaches are reviewed here, to outline those that are relevant for the purpose of this dissertation.

2.1 Presence as realism

As can be understood from the word realism, this approach centers on that technology should produce very accurate representations of objects and events. However, realism in this case only refers to the potential perceptual realism and not the social realism of what is mediated. In other words, this approach uses the real word as a scale on which the medium is assessed and it is often seen in various presence questionnaires in the form of questions like *"How real did the experience feel?"*. As it will be described in details later, different experiments to evaluate the realism of the proposed simulations have been proposed.

2.2 Presence as transportation

Transportation can be divided into 3 main categories: "You are there", "It is here" and "We are together". Common for all categories is that some sort of transportation is on-going; either of the subject or of the mediated environment. This concept of transportation in presence is not only common for VR it is used in e.g., virtual tours on the Internet and is a common symbol of daydreaming and reading books.

2.3 Presence as immersion

The theory of presence as immersion takes advantage of the dualism of the concepts. It focuses on that active use of technology might be productive as psychological benefits can be made from actually enveloping the user in physical based technology, such as Head Mounted Displays, data-gloves and sensory input devices. As such the concept takes on ideas similar to rock-concerts: If there are really big speakers, it must be really loud and therefore people feel it more enjoyable.

2.4   Presence as social richness

Researchers in communication view presence from factors such as warmth, sociable, sensitive or intimacy with the purpose of heighten the possible levels of interactions between people. The theories of social presence on media richness seek to maximize the satisfaction and efficiency whereby tasks in an organization are done with a quicker speed (and thereby with higher satisfaction of management and stockholders). To create such an atmosphere of a social presence the medium used must be able to portray the participant's non-verbal cues of communication. The approach thereby outlines that for a meaningful communication to happen over long-distances, one has to also be able to mediate bodily features such as postures, eye-contact, facial expression and gestures. When the medium does allow for the transmission of such factors a higher degree of social richness will be readily available resulting in a heightened level of overall intimacy. Since in this dissertation interactions between people are not investigated, but only the interaction between one person and the virtual environment he or she is exposed to, the concept of presence as social richness will not be further elaborated.

Common for such definitions are that they share the view that while people are situated in virtual environments, active cognitive processes are ongoing. Even more so, it has been stated that if presence can be defined as a feeling of *"being there"*, then it must also be agreed that this feeling must be the result of an ongoing individual evaluation of the incoming sensory data [10]. However, to me this argument (and the consequences that inherently must follow it) seems to be questionable when the nature of the human biological system is examined. Even though the human perceptual system does perform enormous amounts of processing of data, it would seem questionable that evolution would have forced us into an ongoing process of evaluating stimuli as being real or virtual, the simple amount of processing and data would lead to redundancy; and the shear concept of such an active check of reality (on/off) seems to be a bit overdone since virtual reality has only existed in the last 20 years; a rather minimal amount of time compared to the evolution of mankind.

When the different definitions of presence are analyzed, other aspects about this field of research are revealed. As presence research currently is performed merging different fields, it can be considered as an interdisciplinary research field.

First and foremost it must be understood that presence-research does not have one single common objective. While [13] is mostly interested in HCI related matters concerning VE, it does provide a helpful definition of subtopics that could categorize the presence-community. (a.) human performance efficiency in virtual worlds (b.) health and safety issues, and (c.) potential social implications of VE technology. While such topics can overlap it should prove clear to the reader that the categorization is especially significant when compared to the various fields of studies, such as Computer Science (a) psychology and medicine (b) and psychology and sociological studies (c).

Having such categorization in mind can prove helpful when reviewing the definitions of presence. For example a definition as *the perceptual illusion of non-mediation* may seem very attractive in psychological terms and cognitive studies. While it may have the positive effect of drawing attention away from the technology itself, it seems to be unrealistically ahead of its time, as subjects at present when situated in Virtual Environments have to be introduced to wires, cables and other necessary equipment. Such channels for data streams are necessary, and to me it seems a bit ambiguous to talk about non-mediation when being so physically dependent on for example wires. While the definition itself is intriguing, studies within this field are often based on self-evaluations (as opposed to recorded physical data) and maybe the definition should be viewed in connection to this fact. Opposed to this, a definition of presence in the term of suspending disbeliefs could be seen as very pro-technological, when it is regarded that the feeling of presence is directly connected to technology's ability to sustain and elevate such a feeling.

Further definitions of presence can be found in the field of telepresence. This field of studies is concerned with mediated experiences of localities, and as such addresses presence as transportation. Defining telepresence is multifaceted; when inspecting a definition as the perception of presence within a physical remote or simulated site [10]

both psychological and pro-technological aspects seem to be covered. This definition does not deny that active cognitive processes are on-going while also asserting that what VR does offer is simulation based.

To me the definition offered by telepresence seems to be most appropriate. However, the definition does not offer any true definition of how such perception of presence should be attained, especially defining what the determinants of presence are. While the other definitions mentioned above also go a long way to avoid such clarification, it is my feeling that presence could be defined as a feeling comparable to being there, that may be obtained when a sufficient amount of interrelated multi-modal input is introduced to the subject.

3

Measuring presence

Hand in hand with development and research within the areas of Virtual Reality and Virtual Environments a rather large number of projects of measurements have been set up. With the requirement of trying to reach a definition of what presence is, how it can be defined and how it should be measured, several studies have been made on how content in questionnaires should be formed when carrying out presence research. However, beyond measuring the perceived level of presence, other factors may also be measured in order for one to assess the technology and its level of delivering immersion.

When measuring presence it is easy to direct questions on the topic itself, and as such regard it as a uni-dimensional construct. Questions in such regards will often be mainly concerned on the feeling of being present in virtual environments which by far

and large can be fine. However, at the very base of presence-research a discussion on the approach to perform this research is situated. While presence may be measured as an item on its own, a more holistic ideology may also be taken in approaching the topic; viewing presence as a sum of a multifaceted experience, carrying many different characteristics and phenomena that together creates a feeling of being there. One could argue that many different domains may be represented within the feeling of Presence, e.g. the spatial domain (visual) temporal-spatial domain (sound) and cognitive domain. Traits and characteristics that have a suggested impact are quality, predictability, awareness and others. Furthermore, for revealing biased answers and get a more evolved understanding of the experience of the subject these traits should be tested. For an excellent review of discussions on which characteristics that may be effective please see [11].

An appropriate start is to clearly define two factors: media form and media content [10, 11]. Within a given questionnaire clear references to the media form should not be given. In other words the questionnaire should be considered being useful if the intended media experience was transferred from one media-solution to another; e.g., from a CAVE-installation to a Head Mounted Display. The immediate contribution of following this rule is that the subject will be more inclined to concentrate on recalling the feelings of *"being there"*. So rather than thinking that he is a part of a technological study, the subject will recall his level of attention and involvement; traits often associated with presence [10]. Furthermore a questionnaire that is developed in a non-media typical way will be transferable to comparative studies using other media-platforms. As a third factor keeping questionnaires free from media-related questions can contribute in ensuring that presence is actually measured and not the subject's preference of technical advantages of a media-form.

While intentions of creating questionnaire that are independent from the media-types they investigate have very clear advantages, it is the belief of this author that a truly independent questionnaire type will be very hard to develop. Authors of the ITC Sense of Presence Inventory [10] use this factor as a means of rating questionnaires, which seems too ambitious. With a fast development within the field of VR-

technology, it seems obvious that some current and future media-forms will hold inherent advantages compared to others. While studies show that increasing performance within one modality does not show direct correspondence with measurable immersion [11] some media-types will use different constructs of multimodal environments. It would seem appropriate that research is done in finding proper mathematical factors to calculate results so that media-forms can be compared.

In creating questions care should be taken not to investigate for several factors within the same question [10] . An often used method to avoid this is to formulate the question and answer form after the Likert-scale which will ensure that only 1 factor is being answered by the subject. The Likert-scale [14] was originally designed to enable respondents to specify their level of disagreement or approval of statements. Such statements should be defined as either favorable or unfavorable (sometimes referred to as "'bi-polar emotional response"') toward the topics of the studies. The questionnaire should take special care that factors of Involvement and Immersion are divided, as these seem to have a clear connection [11]. As involvement is increased the level of immersion will often rise, as the attention of the subject is more focused on the virtual environment. For studies that involve several scenarios this might have an effect on results obtained, and a procedure for leveling the amount of involvement might be necessary unless the direct objective of the study is to find the effect of involvement as corresponding to for example augmentations.

Options of responses within the questionnaire type should aim at being identical across questions. If efforts are made to keep available response-options consistent the questionnaire will seem more user-friendly and answering will not cause the subject to have to change the state of mind. When using a Likert-scale (e.g. 1-5.) this could render the questionnaire counterproductive if the next question was answered using a scale in percent.

Questions in various presence questionnaires seem to follow variations of either (a.) the Likert-scale (1-5) or questions that offer a similar approach using percentages. Questions are often formulated [15] like:

1. *How compelling was your feeling of being present in a virtual world?*

2. *How aware were you of the real world while you were navigating the virtual world?*

Even though questionnaires are widely used in presence evaluation, it is not my opinion that they are designed for performing continuous measurements, due to the limitations of human memory. As the participants have to evaluate retrospectively (post-test), these features or events have to be somehow memorable in order for the participants to remember the particular instance and how the experience was during that event. Several alternative presence measurement methods already exist. Even though not all of them produce moment-to-moment results, they all seek to measure and log presence at points during mediation of the test material. One of them is the use of psycho-physiological measurements (measures of heartbeat, muscular responses, etc., as an indicator of presence). Applying physical reactions of the body as a measure of presence is justified by the theory that if one is present in a virtual environment the body of the person will react to stimuli in the virtual environment in the same way as it would in the real world [16]. A disadvantage of this type of measurement is generally that the equipment can pick up low signal-to-noise ratios.

Another suggested way of measuring presence is through behavioral measurements, namely observations of a person's behavior while being exposed to the test material. As with psycho-physical studies, these measurements rely on the idea that if a person feels present in an environment, the person will behave in the same manner in the virtual environment as in the real world, for instance by leaning in the counter-direction of a mediated movement [17]. A benefit of behavioral measurements is that behavior is often triggered spontaneously, without the test person having much control of it. However, a significant downside of this method is that the observer can misinterpret the behavior and that the test has to be specifically designed to a situation where a behavioral response can be provoked from the test persons, which is not necessarily the case for any feature or narrative in the environment which developers want to evaluate.

The usage of breaks in presence [18] also seems interesting as it can provide hints

regarding what elements, in a mediated environment, are likely to induce breaks in presence during a test session [19]. This way of measuring entails that the participant reports every time his or her attention is brought back to reality after the participant has been feeling present in the presented material. The method relies on the test participants' ability to judge whenever their attention has returned from feeling presence. A drawback of this method is that it is not sensitive to different levels of presence, and it does not take into account that a person can be partially focused on both the mediated and the real environment. Another issue is that it requires careful instruction of the test participants beforehand in order for them to be able to report whenever the feeling of presence is broken. Besides breaks in presence other attention-based measuring methods have emerged in an attempt to measure presence continuously throughout an experiment. One variation is the introduction of a secondary task while having the participant performing in a VR experience.

In [20] test-participants were exposed to a monitor displaying a movie with a detailed narrative, while the participant at the same time attended a VR experience. The assumption behind this method is that the amount of presence felt in a space corresponds to the amount of attention directed toward the mediated environment. By having the participants recalling events, which had occurred in either the movie or the VR experience, the attention directed toward each environment during the experiment could be measured. Thereby it could be determined at which times the participant had been more present in respectively the VR environment or the movie. Like breaks in presence, this method lets the developers know whether the participant has felt presence in their media product or not. However, with this method, it is still not possible to say at which level the participant has been feeling presence. Another point is that the questioning still takes place post-test (like questionnaires), and therefore the results still rely on the participants' ability to recall information.

A third attention-based method is the use of reaction time to a secondary task as a measure for presence [16]. In a typical secondary-task reaction time (STRT) experiment, a test subject performs a task (e.g., plays a video game), and at some point, a signal (e.g., a tone) is used to alert the test subject (hence breaks presence). The time

difference is then noted from when the signal was emitted to the time where the test subject responded. This time difference is then considered to be a measure of presence [21]. The underlying assumption behind the STRT method is that as presence increases, more attention (and therefore more mental processing capacity) will be put in the performed task. If more capacity is put into performing the primary task, fewer resources remain for the secondary task (if a limited capacity is assumed), which, in the case of reacting on the signal, could result in a longer reaction time [21]. Reaction time can on the other hand also be affected by a set of other factors. Reported factors are, for instance, muscular tension [22], age [23–25], gender [25, 26] , fatigue [24] and the breathing cycle [27].

Experiments conducted with the purpose of testing the STRT method as an indication of presence have generally returned mixed results [28–30]. However, this method allowed to measure presence while subjects interacted with the media product they were exposed to. This is an advantage compared to the use of alternative techniques for measuring presence such as questionnaires, where subjects are asked to remember an environment they have previously experiences, and evaluate their sense of presence based mostly on their recollections of the experience. Moreover, since the game developed included different tasks to be performed, it was felt appropriate to evaluate whether interrupting such tasks could have an effect on sense of presence. This is the reason why I adopted the STRT method and adapted it for my purposes, as described in the following in my published papers.

4

Presence and sound

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 [3, 4, 31, 32]. Although the visual modality has clearly seen the majority of research efforts in virtual reality research, some important advantages of the auditory system should be taken into consideration when designing immersive virtual environments. First of all, while using vision only a fraction of an environment is perceivable at a time, this is not the case with audition. In [32] is is observed that this limitation of the visual system produced a decrease of the overall realism of a simulation. Moreover, with audition it is possible to sense both direct sound and reflections from all directions in space, and this can happen without turning our heads. This ability of the auditory system allows to get an impression of geometry and size of the environment [32]. Another important characteristic of the auditory system is its higher temporal resolution,

higher compared to both sense of vision and touch [33]. Moreover, the auditory system cannot be turned off, as the visual system can. According to [31], this is a crucial condition which might be determinant for achieving a full sense of presence.

Most of the research relating sound and presence, in fact, has examined the role of sound versus non sound and the importance of spatial qualities of the auditory feedback [32, 34]. Other prior research has addressed issues related to the addition of auditory cues in virtual environments, and whether such cues may lead to a measurable enhancement of immersion in such environments. Most prior work in this area has focused on sound delivery methods [6, 35] and 3D sound [34]. Even less is known about the role of interactive auditory feedback in enhancing sense of presence.

An often reoccurring theme in research related to presence induced by the auditory modality is that of congruency between the auditory and the visual display [36–39]. Consistency may be expressed in terms of the similarity between rendered visual and auditory spatial qualities [40], the methods of presentation of these qualities [34], the degree of auditory-visual co-occurrence of events [38, 41] and the expectation of auditory events given by the visual stimulus [39].

A smaller amount of research has been concerned with investigating the quality and quantity of the auditory feedback and its relation to presence [38, 39]. In particular, in [39] the approach of ecological perception is taken, and it is proposed that expectation and discrimination are two possibly presence related factors. Expectation is defining as the extent to which a person expects to hear a specific sound in a particular place, and discrimination is the extent to which a sound will help to uniquely identify a particular place. The result from their studies suggested that what people expect to hear in certain real life situation can be significantly different from what they actually hear. Furthermore, when a certain type of expectation is 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.

Another interesting direction of research has been concerned with understanding the effect of lack of auditory cues in providing sense of presence. In [42], a series of experiments is carried out where participants were fitted with earplugs and instructed to carry out some everyday tasks during twenty minutes. Afterwards, the participants were requested to account for their experience and to complete a questionnaire comprising presence-related items. Overall, support was found for the notion of background sounds being important for the sensation of being part of the environment, termed in [42] as environmentally anchored presence.

Paradoxically, the study also suggested that due to the lack of auditory cues, the participants also had an increased attention to what was happening, they felt as being more aware of the situation than normally. Finally, the use of the earplugs also resulted in a situation where participants had a heightened awareness of self, in that they could better hear their own bodily sounds and which in turn contributed to the sensation of un-connectedness to the surround (i.e. , less environmentally anchored presence). In addition to stressing the importance of the auditory background in VE, auditory self-representation can be detrimental to an overall sense of presence [42].

In [43], a study is described in which participants assessed their sense of presence obtained with binaural recordings and recorded video sequences presented on a 50-inch display. Taking a broader perspective, [43] performed experiments 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, i.e., footsteps, vehicles, doors etc. In their study, they found that especially two factors had high positive correlation with sensed presence: sound information and sound localization. The results also showed an interesting auditory visual integration effect: when the sound was matched with a visual sequence where the sound source was actually visible, presence ratings were higher. This implies that there are two important considerations when designing sound for VEs, one being 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.

As suggested in [43], the spatial dimension of sound is not the only auditory determinant of presence. In support of this, [34] found no significant effect of adding three extra channels of sound in their experiment using a audiovisual rally car sequence. On the other hand, they found that enhancing the bass content and sound pressure level (SPL) increased presence ratings. In a similar vein, [41] found that presence ratings increased with reproduced SPL for conditions without any visual stimulus, but that sensed presence in general was highest for realistic SPLs when visual stimulus was presented simultaneously with auditory stimuli. In audiovisual conditions with a inside moving car stimulus however, presence was highest for the highest SPL, which was explained by the fact that increased SPL may have compensated for the lack of vibro-tactile stimulation.

A study presented in [35] compared conditions in virtual reality with no sound, surround sound, headphone reproduction and headphone plus low frequency sound reproduction via subwoofer. Both questionnaires and psycho-physiological measures (temperature, galvanic skin response) were used to access affective responses and presence. All sound conditions significantly increased presence, but only surround sound resulted in significant changes in physiological response followed by a marginal trend for headphones plus subwoofer condition. Interestingly, questionnaires did not show such discrimination between sound reproduction techniques confirming unreliability of this measure when used alone.

In [44] it is suggested that proper relations between auditory and visual spaciousness is needed to achieve a high sense of presence. In their experiment, a visual model was combined with two different acoustic models; one corresponding to the visual model and one of approximately half the size of the visual model. The models were represented by means of a CAVE-like virtual display and a multichannel sound system, and used in a experiment where participants rated their experience in terms of presence after performing a simple task in the VE. Although some indications were found sup-

porting that the auditory-visually matched condition was rated as being the most presence inducing one, the results were not as strong as predicted. An explanation to these findings was that, as visual distances and sizes often are underestimated in VEs [45], it was likely that neither the proper sized acoustic model, nor the wrong sized acoustic model corresponded to the visual model from a perceptual point of view. Thus, a better understanding of how visual spaciousness or room size is perceived would be needed to perform further studies on this topic. Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in VEs [46–48].

Self-generated sounds have been often used as enhancements to VEs and first-person 3D computer games – particularly in the form of footstep sounds accompanying self-motion or the presence of other virtual humans. It has been suggested that also the degree of consistency within modalities would affect presence [4]. In the auditory domain, an example of an inconsistent stimulus could be a combination of sounds normally associated with different contexts (e.g., typical outdoor sounds combined with indoor sounds). Another type of within-modality inconsistency could be produced by spatializing a sound with a motion trajectory not related to that particular sound. In our example with the passing car above, this situation could occur e.g. if the sound of a car driving at slow speed is convolved with an HRTF trajectory corresponding to a high-speed passage. However, although these types of inconsistencies intuitively seem to be detrimental for the presence sensation, there is, to the best of my knowledge, little evidence to support this notion. In sum, it can be seen that a general understanding of various auditory display factors' contribution to the sense of presence begins to emerge. It is however clear that the findings presented above need further corroboration with different content and methodologies.

This is one of the reasons why I felt the need to investigate further the issue of auditory feedback and presence, as outlined in the first set of papers presented in this dissertation.

5
Contributions

In this dissertation, new evaluation techniques for multimodal environments based on presence studies are proposed. Moreover, the connection between multimodal interaction and presence, when applied to different media, is analyzed.

The dissertation can be seen as divided in two sets of papers. The first set of papers investigates how presence research can inform the field of sonic interaction design. Specifically, presence and sonic interaction in a walking context are investigated. One of the motivations for choosing to work with walking sounds is that they have long played an important role in audiovisual media. In film, footsteps are acknowledged for their ability to signify unseen actions, to lend a sense of movement to an otherwise static scene, and to modulate the perception of visible activities. In his seminal work on film sound, Chion writes of footsteps sounds as being rich in what he refers to as materializing sound indices, those features that can lend concreteness and materiality

to what is on-screen, or contrarily, make it seem abstracted and unreal [49]. Contact interactions between our feet and the ground play important roles in generating information salient to locomotion control and planning in natural environments, and to the understanding of structures and events in them. Although much of this information is communicated as sound, the latter has been relatively neglected in past research related to walking in human computer interaction. Consequently, a better understanding of the perception of walking sounds, and the way they may be rendered and displayed, is needed in order for new and existing human-computer interfaces to effectively make use of these channels. Such developments may hold potential to advance the state of the art in areas such as wearable computers, intelligent environments, and virtual reality. In researching multimodal interfaces based on walking, an interdisciplinary approach is essential. A complete research plan on walking for multimodal interfaces, in fact, must include understanding of physics of walking, basic results on psychophysical aspects, integration in multimodal installations. Such have been investigated in the first set of papers.

The second set of papers investigates how presence research can be adapted to evaluate user's experience in an interactive and non-interactive scenario, i.e., while playing a computer game and watching a movie. Both sets of papers investigate to what degree the design of alternative interfaces helps presence research.

The contributions of this dissertation can be described as follows. First of all, new evaluation techniques for measuring presence in multimodal environments are proposed. The focus has been placed mostly on environments where presence research has not been widely investigated yet, such as custom made interfaces which control synthesized sounds, and computer games. The first set of papers examines the role of evaluation techniques and presence research in sonic interaction design. The field of sonic interaction design has been merely concerned with the development of novel sound effects algorithms which work in real-time and are controlled by alternative interfaces, either commercially available or custom made. In interaction design, the availability of physical computing resources has put the construction of interactive sonic objects among the favorite activities of many practitioners. The emergence of

the discipline of Sonic Interaction Design is facilitated by the increasing possibilities offered by sensors and actuators technology. Complex body gestures can nowadays be captured, and tightly coupled to interactive sounds. On the other hand, evaluation techniques for sonic interactive products are still lacking [2]. In this dissertation, I propose evaluation techniques for interactive products with a salient sonic content which are based on presence research. We propose the combination of qualitative and quantitative evaluation techniques, combining measurements of users' actions with presence questionnaires and interviews.

Among scholars in perception and cognition there has been a shift in attention, from the human as a receiver of auditory stimuli, to the perception-action loops that are mediated by acoustic signals [50].

The main focus is on ego-sounds produced by the action of walking. A large study whose goal is to evaluate the role of interactive auditory feedback in enhancing motion of users in a virtual environment is described, and the relationship between auditory feedback and presence is investigated.

As described previously in the chapter on presence and sound, most of the previous research investigating the connection between sound and presence has focused on the role of presence or absence of sound or the role of different sound delivery methods or sound rendering [44]. Few studies have examined how the content of the sound affects presence.

Moreover, to my knowledge, nobody has previously investigated how sounds enhance motion of subjects in a virtual environment. Most of the studies combining sound and motion have been performed in the field of music performance [51]. Here, I focus on virtual environments where everyday sounds and a photorealistic visual feedback is provided.

Furthermore, to my knowledge nobody has evaluated new technologies such as custom made interfaces which drive interactive sound effects from a presence perspective. To my knowledge, the effect of such self-generated sounds on users' sense of presence had not been investigated prior to the author's research in this area. The combination of physics-based rendering of walking sounds with contact-based sensing also

appears to be novel. So this thesis brings together the field of presence research with the field of sonic interaction design.

The studies described in papers I and V show how the sound produced by the subjects' own footsteps is not sufficient to enhance the subjects' motion in the environment. What makes the all experience more engaging is the combination of interactive footsteps with the soundscape of the specific place.

This observation motivated us to run a workshop in Porto during the Summer 2009, as part of the Sound and Music Computing Summer School. The goal of this workshop, described in paper III, was to combine the two interactive footsteps controller developed as part of the Natural Interactive walking project with the soundscape of the city of Porto. Specifically, I encouraged students to find some interesting locations in Porto, record them and create a soundscape which represented the city. In the laboratory, such soundscape was combined with the interactive footsteps controller. This resulted in an installation where subjects were feeling as if they were walking around landmark locations of Porto. Although not formally tested in this context, the visitors of the installations commented that the experience of using the footsteps controller was greatly enhanced when the soundscape was provided. This confirms the results formally tested in the experiments described in papers I and V.

Papers I to V have the following progression in chronological order from the most to the least recent. For clarity I describe the research progression from the oldest to the newest paper. Paper V, presented at CHI 2008, presents results on a large study which examines how motion of subjects in enhanced in a photorealistic virtual environment. Paper IV describes different issues which need to be taken into account when designing walking interactions. The paper is co-authored by different members of the European project Natural Interactive Walking (www.niwproject.eu). Expertises of the authors covered in the paper range from design of floors which provide haptic feedback, design of shoes enhanced with sensors, sonic feedback to simulate sensation of walking on different surfaces as well as evaluation of the systems. I mostly contributed to the evaluation part in the paper, which describes both basic psycho-physical experiments, experiments which evaluate the affective content of footsteps sounds and presence and

immersion studies.

As described before, paper III explores the use of the developed walking interfaces and sound synthesis algorithms in the context of the city of Porto. A workshop is described where interactive footsteps driven by the users are combined with the soundscape of the city of Porto, in such a way to be able to re-create in a laboratory the sensation of walking in the city. After improving the design of the interactive footsteps controller, it was felt the need of performing basic psycho-physical experiments to validate its usability, before entering advanced presence and immersion studies: this is why the experiments described in Paper II were performed.

Paper II explores basic evaluation of the interactive footsteps synthesizer described in Paper III. The paper proposes and justifies the design of different alternatives to synthesize and control footsteps. Such design is tested in the paper. The importance of finding a tradeoff between complexity of the interface/interaction and ability to provide sense of presence is discussed. This is one of the reasons why the research started with a simple interface (shoe with one sensor) and then increased it in complexity. Obviously a tradeoff exists between presence and the complexity of the interface, since complexity also obviously implies more cables, latency, etc.

The papers therefore present an evolution in the development of the technology used. Starting with shoes with only one pressure sensors (paper V) and their design was then improved as described in papers III and IV. Paper II describes basic evaluation techniques of the technology developed. Finally, paper I reports on more in depth presence and immersion studies.

Papers VI, VII and VIII propose a novel methodology to evaluate presence in mediated environments, which is called adjustable distraction method (AD) from now on. Papers VI and VII show how the method has been applied to a computer game and a movie respectively, while paper VIII describes a larger study where some of the experiments described in the previous paper has been redone, mainly for the purpose of validating the previous results. Two main experiments were run: the first experi-

ment consisted of using a visual distracter to evaluate presence as immersion [12] in a computer game. The experiment also evaluated the intrusiveness of the AD method.

The second experiment investigated whether heart rate measurements, intensity ratings and the results of the AD method with vibration as the distraction signal were comparable when test participants watched a movie clip. The main idea in proposing the AD method is to find an alternative to the use of the techniques to measure presence that have been described in the state of the art section, together with their advantages and disadvantages. The starting point of the AD method is the previously described attention-based secondary-task method [20]. A number of test participants are here given a simple secondary task to perform, namely to react whenever they are exposed to a primitive stimulus.

A primitive stimulus is not meant to be another mediated virtual experience such as it is the case in the secondary-task method. Instead, the stimulus should be regarded as a distraction, as in the STRT method, which immediately reminds the participant about the fact that he or she in reality is participating in an experiment and has to respond now. The stimulus used could be any perceivable signal (e.g., a visual, an audible or even a tactile signal). The important factor is that the signal can vary in strength, from undetectable to unavoidable.

As it is the case with the breaks in presence method, in the AD method the participants are asked to react whenever presence is broken. However, in contrast to the breaks in presence method, the participants do not need a careful explanation of the presence concept in order to respond. The proposed method also has great resemblance to the STRT method in the sense that they are both measuring sustained attention. As such, the STRT method has all the advantages of the AD method (e.g., temporal measures, objective, sensitive to different levels, not retrospective).

However, the AD method does not rely on time as a measure (but on the amount of distraction) and therefore all factors potentially leading to either increased or decreased reaction times (e.g., age, gender and recent muscular activity) can be disregarded in connection with the use of the AD method. A mechanism to present to test subjects an adjusted stimulus is also proposed.

In all the papers presented in this dissertation, when a presence study is conducted the presence questionnaire named Swedish Viewer User Presence Questionnaire (SVUP) [52] was adopted.

As described in Chapter 3, several questionnaires have been proposed in the literature to measure presence. However, I believe that the SVUP questionnaire well served my purpose and addressed issues which I was interested in measuring. The questionnaire was adapted by removing items which were irrelevant in the current investigation, such as questions concerning social presence. Moreover, common questions from validated presence questionnaires were also used and adapted to our context.

It is my opinion that it is novel to merge the fields on interface design and presence research. Especially concerning the field of sonic interaction design, interfaces which are used to control sonic feedback have not yet been investigated from a presence perspective.

Furthermore, that it is extremely important to put these interfaces in a context. This is the reason why I proposed and run the workshop described in Paper III, where the interfaces designed have been placed in the context of the city of Porto.

Moreover, it is extremely important to evaluate novel technologies not only from a higher level presence and immersion perspective, but also from basic psychophysical perspective. This is the reason why the experiments described in Paper II were conducted. Finally, it is important to find a tradeoff between complexity of the interface/interaction and ability to provide sense of presence. This is why the research started with a simple interface (shoe with one sensor) and then increased it in complexity.

The same can be said concerning the second set of paper, where first investigations have been conducted with a simple visual feedback, and then other sensorial modalities such as tactile feedback have been introduced.

6

Description of the submitted papers

6.1 Paper I - R. Nordahl. Evaluating environmental sounds from a presence perspective for virtual reality applications, EURASIP Journal of Audio, Speech and Music Processing, Special issue on Environmental Sound Synthesis, Processing and Retrieval, 2010.

This paper describes the design of a multimodal environment to provide users with IBR (Image Base Rendering) visual information and various levels of auditory information: ego-motion (footsteps), static sounds, dynamic sounds and music. It starts by describing the need for high quality audio in multimodal virtual environments and the notion of presence in such environments. A review of the state of the art research on presence and auditory feedback is presented.

As described in Chapter 4, research on presence and auditory feedback has focused mostly on sound delivery methods and their role in creating sense of presence. Some work has also investigated the role of sound content in creating sense of presence [34, 38,39]. However, to our knowledge nobody has previously investigated how interactive auditory feedback affects presence in virtual environments.

The paper presents a large study whose goal is to investigate the role of different sound delivery methods in enhancing sense of motion in virtual environments, and investigates to what degree such environments create a sense of presence. The research was first motivated by previous investigations where subjects were asked to visit a virtual place where visual feedback was rendered by using IBR techniques, displayed by using an Head Mounted Displays.

In such simulated environments, where scenes were extremely static (subjects could control only the point of view of the environment, but there was no actual action or moving objects in the environment) it was observed that users moved very little, and were not motivated to explore the environment. To overcome this limitation, other forms of feedback which could vary over time were investigated, with the purpose of creating a more immersive experience. I hypothesized that auditory feedback was an ideal time-varying feedback which could allow subject to engage and move in the environment.

To investigate not only the presence or absence of auditory feedback, but also its quality, several soundscapes were designed. Such soundscapes, six in total, ranged from a piece of music and environmental sounds, both static and dynamic. The more complete solution, which was called *Full* in the paper, consisted of rendering soundscapes in 3D with moving sound sources, as well as rendering interactively footsteps sounds. Such sounds were obtained by creating some custom made sandals enhanced with sensors. Such sensors detected the act of users walking in the environment, and triggered a real-time footsteps synthesizer which simulated the sound of walking in the virtual surface depicted.

In the six different conditions to which the subjects were exposed, the body mo-

tion was tracked. Results show that interactive footsteps sounds, together with the corresponding soundscape, significantly enhanced the motion of subjects in the environment. Rather surprisingly, the sense of presence, measured by using a presence questionnaire, was not significantly enhanced in the conditions with interactive auditory feedback and soundscape combined. This can be due to the difficulties of subjects to interpret the questions in the presence questionnaire.

6.2 Paper II - R. Nordahl, S. Serafin and L. Turchet. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. Proc. IEEE Virtual Reality, 2010.

In this paper, a study whose goal is to evaluate ability of subjects to recognize synthetic footsteps sounds in both an interactive and an off-line context is presented. Subjects' perception in terms of quality of the developed system and realism of the simulation is also investigated through an psychophysical test. The ultimate goal is to develop a high quality footsteps synthesizer where users can wear their own footwear and which is efficient enough yet of high quality to be used in virtual reality applications.

While walking, the net force F exerted by the foot against the ground can be represented by a time varying spectrum $F(a, t)$, having components tangential and normal to the ground surface, where a denotes angular frequency and t is time. The term Ground Reaction Force (GRF) is often used to refer to the low frequency information in F, below about 300 Hz. The GRF is essentially responsible for the center of mass movement of the individual. It is approximately independent of footwear type, but varies between individuals or walking styles [53].

Higher-frequencies components of $F(a, t)$ can be attributed to fast impacts between heel or toe and ground, sliding friction and contact variations between the shoe and the ground [54]. Unlike the GRF, these components can depend on footwear, on ground surface shape and material properties. They give rise to remote signatures in the form of airborne acoustic signals, seismic vibrations of the ground, and vibrations transmitted through the shoe to the foot, which have been studied in prior literature on

acoustic [54, 55] and vibrational [56] signatures of human walking. These signals vary with the local material and spatial structure of the ground and with the temporal and spatial profile of interactions between the foot of the walker and the ground surface.

The system proposed in this paper is based on microphones, which capture the sound of a person walking and from such sound extract temporal information concerning the step. More specifically, the GRF is extracted from the acoustic waveform. The goal is to extract the users' contribution to the step, i.e., the GRF, and simulate the floor's contribution, i.e., the material on which the person is walking. The ultimate goal is to create the sensation of walking on a different surface than the one subjects are actually walking upon.

Experiments on recognition show that subjects are better in recognizing sounds created with the interactive system developed compared when listening to prerecorded footsteps. A Chi-square analysis, however, shows that such recognition is significant only for two of the simulated surfaces, namely dry leaves and metal. This means that no real conclusions can be drawn on the better ability of subjects to recognize sounds in an interactive rather than off-line scenario. However, the developed system is promising since results show a high recognition rate. Moreover, informal interviews performed after the recognition test showed that subjects appreciated the quality of the system and the fidelity of interaction. One advantage of the system used in this research, as compared to other systems used in previous research (such as the shoes enhanced with sensors) is the fact that it can be used by subjects wearing their own footwear. However, one disadvantage is the fact that the microphones capture any environmental sound, preventing such system to be used with other external sound sources rather than footsteps, or using it with soundscapes which would add a context to the virtual scenario, and are essential in experiments testing sense of presence.

6.3 Paper III - R. Nordahl, S. Serafin and F. Fontana. Exploring sonic interaction design and presence: Natural Interactive Walking in Porto. Proceedings of Presence 2009.

In this paper, two different interfaces to control synthesize footsteps are compared, their role in simulating the sensation of walking in a specific place, in this case the city of Porto is described.

The first interface, developed at the University of Verona, is a pair of shoes enhanced with sensors. Specifically, the shoes are a pair of sandals with two force sensitive resistors, one for each shoe. The shoes are used to control synthesized sounds of footsteps on different surfaces. The second interface consists of four microphones which are placed on top of a solid surface in a square configuration. The microphones detect footsteps of a person walking on the surface. An algorithm extracts in real-time the ground reaction force from real footsteps, and uses it to control the temporal variations of synthesized sounds of footsteps on different surfaces.

The two developed interfaces are analyzed in terms of their ability to be used in a multimodal virtual environment, and advantages and disadvantages of the two devices are discussed. The ultimate goal is to design an interface which can be utilized in multimodal environments (complemented with haptic and visual feedback) where presence studies can be performed.

The shoes enhanced with sensors have the advantage of being portable. Their main advantage, however, is the fact that they can be used also when other sounds are present in the environment, being those soundscapes synthetically generated or unwanted sounds existing in the environment where the installation is setup. However, the main disadvantage of this setup is the fact that users need to wear custom made shoes. Moreover, the setup is wearable, meaning that all the processing happens on a laptop placed on a backpack carried by the users. This fact has the advantage that users can walk in an unlimited space, but the disadvantage that their way of walking is not natural, likely preventing sense of presence.

The main advantage of the configuration with microphones is the ability of subjects

to wear their own footwear. This is an advantage from the point of view of naturalness of the interface and the interaction. However, the system does not work when other sounds either than footsteps are present in the environment. The conclusion is that neither of the systems is optimal to perform presence studies in a multimodal context, and new solutions are been currently investigated.

This paper was presented at the Presence 2009 conference which took place in Los Angeles in November 2009. The audience at the conference admired especially the combination of custom made devices with studies examining sense of presence. Traditionally, the presence community in fact performs their studies using devices which are available in the market.

6.4 Paper IV - R. Nordahl et al. Sound design and perception in walking interactions. International Journal of Human-Computer Studies, 2009.

This paper is a collaboration between several authors. Most of the authors are partners in the European project Natural Interactive Walking. The main goal of the paper is to outline all the different issues which are important to be researched when designing multimodal interfaces, in this case based on walking. It is beyond the possibility of a single researcher to be able to cover all those competencies in depth; this is why an interdisciplinary collaboration between different skills is desirable.

The paper starts with motivating the importance of a multimodal approach when studying walking. From a sensory standpoint, in addition to vision, the pedestrian receives sound information via the auditory channel, vibrational information via the tactile (touch) sensory receptors in the skin of the feet, and information about ground shape and compliance via the proprioceptive sense (the body's ability to sense the configuration of its limbs in space). Proprioception, vision, and the vestibular (balance) sense are integrated to inform the pedestrian about his motion in space. As can be seen from the forgoing description, walking generates a great deal of multisensory information about the environment. Prior research has emphasized the influence of visual, haptic, vestibular, and proprioceptive information on planning and control of

locomotion over flat surfaces (e.g., [57]).

In two respects, these studies provide a limited account of the complexity of walking in real word environments. Firstly, they have not addressed the range of ground surfaces and materials met outside the lab (e.g., to our knowledge, none has investigated locomotion on gravel before). Secondly, they ignore the information contained in sounds generated by walking on real world surfaces (e.g., acoustic information about the gender of a walker [55]). These limitations are addressed in human perception studies presented in this dissertation. Notably, in VR contexts, when such layers of perceptual information are available, they are likely to contribute to a heightened sense of presence in a virtual environment.

The paper starts by reviewing the literature on walking in virtual environments. It then outlines the different issues which need to be taken into account when designing and evaluating interfaces based on walking. First of all, the paper discusses several interfaces which have been built to track a person walking. In the context of the Natural Interactive Walking project, the participants focused both on haptic floors, by enhancing the interaction of the floor with haptic feedback, and on shoes enhanced with both sensors and actuators. The auditory feedback designed is also discussed, together with the physically based and physically informed algorithms designed to simulate walking interactions.

What the author found particularly valuable when collaborating on this paper was to realize how such a simple everyday task like the act of walking becomes extremely complex when the goal is to faithfully reproduce it in a virtual environment. Although different solutions for haptic, auditory and also visual feedback are proposed, together with methods for evaluating such solutions, an overall evaluation of the advantages and disadvantages of each solution is not provided, and is still an open research issue.

6.5   Paper V - R. Nordahl. Sonic Interaction Design to enhance presence and motion in virtual environments. Proc. CHI 2008.

This paper was presented at the Computer Human Interaction (CHI) 2008 conference, in a special session dedicated to sonic interaction design. The paper describes a large study whose goal is to investigate the role of different forms of auditory feedback in enhancing motion of subjects in a virtual environment.

The motivations behind this work are the results of the BENOGO project, an European project investigating the reproduction of physical places in the virtual world using image based rendering (IBR) techniques. In the BENOGO project, real places were photographed with a machine rotating around itself in 360 degrees. Such places were virtually reconstructed using IBR, in order to give to users the impression of navigating in the place, i.e., being there without going (hence the name BE-NO-GO). One of the peculiarities of this project is the fact that, in order for the IBR technique to succeed, no moving objects need to be present in the environment. This means that, from the visual point of view, the visited environment is extremely static. From the point of view of presence research, i.e., investigating if subjects felt present in the reproduced place, the project was rather unsuccessful, since subjects were feeling extremely bored in visiting such a static environment. The goal of this paper is therefore to provide a temporal dimension to the IBR based simulations provided by the BENOGO project. By using auditory feedback, which is notoriously evolving over time, I hypothesize that subjects will become more interested in visiting the virtual environment. I therefore tracked subjects while visiting the environment, in this case a botanical garden in the city of Prague, and I provided six different kinds of auditory feedback. The auditory feedback ranged from a static soundscape, to a piece of music to interactive shoes enhanced with sensors, which control synthesized footsteps. Results show that the auditory environment where interactive footsteps and 3D sound is rendered significantly enhance the motion of subjects in the virtual environment.

6.6 Paper VI - R. Nordahl and D. Korsgaard. Distraction as a measure of presence: using visual and tactile adjustable distraction as a measure to determine immersive presence of content in mediated environments, Virtual Reality, Springer, November 2009.

This paper presents several studies whose goal is to evaluate the role of distracters in determining presence in two kinds of mediated environments: movies and games. The paper is an elaborated journal version of the research described in papers VII and VIII. Specifically, several studies are presented where the use of the adjustable distracter method is proposed as a way to evaluate presence in mediated environments.

The method has been applied to videogames and movies, and implemented both using visual and tactile feedback. The main assumption behind the method is that presence is as strong as the minimum amount of stimuli required to break it. Therefore the method proposed follows some ideas from the secondary task reaction time technique [21], trying to overcome its limitations.

The main characteristic of an adjustable distracter is that it is a signal which can be perceived, and varies in strength. In the first described experiment, the method was used to evaluate presence in the context of a videogame. In this experiment, the adjustable distracter is visual, and in the form of a circle placed outside the player's region, in a black frame surrounding the game's space. In the second experiment the distracter is tactile, in the form of a signal provided to users when watching a movie.

The first experiment showed that the adjustable distracter method is not intrusive in the context of a game. This makes the method more interesting compared to similar techniques such as the secondary task reaction time. However, a player's experience of intensity while playing a game is not proportional to the minimum amount of visual distraction needed to attract the player's attention, so one of our original hypotheses had to be rejected. The second experiment, where tactile feedback is used, does not show conclusive results on how such feedback can be used to evaluate presence in movies. I believe this might due either to the fact that subjects perceived the tactile

stimuli differently.

6.7   Paper VII - R. Nordahl and D. Korsgaard. On the use of adjustable distraction as a measure to determine sustained attention during movie clips. Proceedings of Presence 2009.

This paper presents research which is a continuation of the one described in Presence 2008. Precisely, the adjustable distracter method was adapted and compared to two other methods for measuring presence, specifically heart rates and subjective intensity rates. The field of research of this paper is examining presence and immersion in movies. In particular, it is the author's interest to investigate whether the adjustable reaction method can overcome the drawbacks which appear in other techniques to measure presence in movies, such as the use of secondary task reaction time (STRT). The adjustable distraction was in this case provided by using a vibration generated by a small speaker. The results of the experiments show similarities between results from questionnaires and heart rate measurements, but the adjustable distraction method reported different results. Great variance was measured in the recorded vibration of the different subjects. This problem might be overcome by having a pre-screening of the subjects, to ensure uniform perceived threshold. Another solution could be to average the results of each subject based on their previously screened perceived threshold.

6.8   Paper VIII - R. Nordahl and D. Korsgaard. On the Use of Presence Measurements to Evaluate Computer Games. Proceedings of Presence 2008.

In this paper it is described a novel methodology to measure user experience in computer games and interactive environments in general. The main contributions of this paper are a novel evaluation technique to measure experience in games. Until recently many or even most of the evaluation methods available to the presence community have been so-called pen-and-pencil-methods. They offer researchers the possibility to allow test subjects to record results, or rather their personal impressions, of the experiences. Drawbacks of such methods are stated in literature and in the paper:

- common issues are the issues of accuracy (how much can be trusted what test subjects report since this is not an actual measurement),

- paper-and-pencil methods records the experiences retrospectively. As such it is not informa-tion from the actual experience that is recorded, but rather the memory of the experience from the individual test subject, that has had time (and is even asked to further process this experience by the mere act of filling out a questionnaire)

- Typical interactive experiences such as computer games are of their nature not truly linear experiences. Therefore each test-subject will have had a, at least slightly, different path through the computer game scenario.

- Test-subjects may be positively or negatively biased (of many different reasons). Issues such as feelings, moods, personal prejudices etc. may all affect the way the test-subject chooses to answer the questionnaire.

In this paper, the goal was to develop a novel alternative that would allow measurements of experiences in computer games. The method is thought as a tool for developers of computer games that need to input during the development period of a product (concerning immersion, engagement and presence). The novel technique proposed was coined adjustable distraction. The paper sought to combine data recorded through a computer application with data from more traditional questionnaires.

The technique proposed consists of using a distracter which becomes increasingly noticeable and asking subjects to press a button when the distracter is noticed (and the immersion in the computer game is thereby broken). Specifically, the screen was divided into two sections, one being a black frame containing a white dot. The dot's colors changes from black to gray and represents the distracters.

A novel adaptive algorithm was also proposed which allows changing the size of dots according to users' responses. This is in order to measure several levels of presence, instead of simply measuring the ON/OFF possibility of the button being there or not. Finally a computer game was build to use as basis for running the experiments.

6.9   Connections between papers

Papers I, II, III, IV and V examine how presence can be adapted to the field of sonic interaction design. Specifically, novel techniques to adapt presence research to the design and evaluation of walking interfaces are proposed.

The papers are organized in anti-chronological order, from the newest published to the oldest. In paper V the design of a photorealistic virtual environment that users can visit by using an head mounted display is described. The environment consists of a reproduction of the botanical garden in Prague. The visual feedback was enhanced by using different combinations of auditory feedback. Results show that interactive auditory feedback significantly enhances the motion of subjects in the environment. Paper IV describes different issues which need to be taken into account when designing walking interactions. The paper is of interdisciplinary nature, and co-authored by different members of the FET Open EU project Natural Interactive Walking. Expertises of the different authors covered in the paper range from the design and implementation of floors which provide haptic feedback, to the design of shoes enhanced with sensors, to sonic feedback to simulate sensation of walking on different surfaces and evaluation of the different systems, both in terms of perception and in terms of presence. The author contributed mostly to the evaluation part in the paper, which includes both basic psychophysical experiments, and which evaluate the affective content of footsteps sounds and presence and immersion studies. Paper III explores the use of the developed walking interfaces and sound synthesis algorithms in the context of the city of Porto. A workshop is described where interactive footsteps driven by the users are combined with the soundscape of the city of Porto, in such a way to be able to recreate in a laboratory the sensation of walking in the city. Paper II explores basic evaluation of the interactive footsteps synthesizer described in Paper III. The paper proposes and justifies the design of different alternatives to synthesize and control footsteps. Such design is tested in the paper. Finally, Paper I investigates in details sense of presence and self-motion in a walking interface, extending the investigations started in Paper V.

All papers have in common the design, implementation of alternative interfaces to simulate walking in virtual reality, and the evaluation of such interfaces from a hu-

man centered perspective. The papers present an evolution in the development of the technology used. Paper V presents studies performed with shoes with only one pressure sensors. An improved design is presented in papers III and IV. Paper II describes basic evaluation techniques of the technology developed, such as psychophysical experiments which assess ability of subjects to recognize the simulated surfaces they are stepping upon. Finally, paper I reports the results of an in depth presence and immersion studies conducted with the technology developed.

The second set of papers examines how presence studies can be adapted to the evaluation of computer games. The papers examine how distracters can be used to evaluate presence in computer games and movies. Paper VIII presents an experiment where a a visual distracter is adopted. A similar technique is also used to evaluate presence in the context of a movie clip, as described in Paper VII. The advantage of using a movie clip as opposed to a videogame is the fact that the linearity of the media allows a better control to the stimuli subjects are exposed to. Finally, paper VI presents a large study which combines studies presented in Papers VII and VIII, and extends them by also using a tactile distracter.

Both sets of papers examine how presence methodology can be adapted to be used in evaluation of novel media technology. In both sets of papers, custom made interfaces are designed to support the investigation of presence. Specifically, in the first set of papers custom made shoes with sensors are designed, while in the second set of paper a vibrator to induce tactile distraction is designed.

48   CHAPTER 6. DESCRIPTION OF THE SUBMITTED PAPERS

7

Conclusions

In this thesis presence research is investigated as a method to evaluate users' experiences in virtual environments.

In particular, presence is applied in disciplines where it was not applied before, such as sonic interaction design, computer games and movies.

The first set of papers explore the relationship between auditory rendering and presence. In particular, I investigated the different sonic elements which are present in an environment, and assess their ability to convey a sense of presence. By combining different kinds of auditory feedback consisting of interactive footsteps sounds created by ego-motion with static soundscapes, it was shown how motion in virtual reality is significantly enhanced when moving sound sources and ego-motion are rendered. These results are demonstrated in Paper I and V submitted with this dissertation.

On the other hand, presence did not appear to be significantly enhanced when com-

bining interactive footsteps with surround sound, as opposed to using a static sound-scapes. I therefore investigated further the role of the combination of soundscapes and interactive sounds in virtual environments. My investigation was justified by the fact than, when exploring a place by walking, two main categories of sounds can be identi-fied: the person's own footsteps and the surrounding soundscapes. During a workshop which I ran during the Summer 2009 in Porto, I investigated whether combining a recorded soundscape of the city together with simulated interactive footsteps driven by subjects enhanced sense of presence. Subjects who tested the footsteps interface with the soundscape reported a higher perceived realism and immersion as opposed to those who tested it without the soundscape.

Overall, I found particularly interesting to realize that a simple everyday task like the act of walking becomes extremely complex when the goal is to faithfully reproduce it in a virtual environment. Different solutions are proposed in this thesis, together with methods to evaluate such solutions. Presence as realism and fidelity of interaction proved to be particularly important when reproducing the act of walking. Subjects, infact, commented on the realism of the simulated footsteps sounds, and the discrep-ancy between auditory and haptic feedback. As a matter of fact, for some subjects appeared as strange to hear an aggregate surface but feel a solid one, i.e., the floor of the laboratory.

Overall I believe that presence research represents a strong method to assess user experiences in multimodal environments. This is why I adopted it also to videogames and movies.

I propose a new evaluation method called adjustable distraction which assumes that presence is as the minimum amount of stimuli required to break it. In the first set of experiments I apply to measure presence in the context of a video-game. In this situation a visual distractor was used, namely a dot of varying place and size outside of the players region of interest. I showed that the adjustable distractor method is not intrusive. Therefore I concluded that this method is more interesting compared to similar techniques used in computer games evaluation, such as the secondary task reaction time. I could not find proportionality between players experience of intensity while playing a game and the minimum amount of visual distraction needed to attract the players' attention.

In a second experiment I investigated whether the adjustable distraction method could work by using tactile feedback instead of visual feedback. The experiments did not show conclusive results on how tactile feedback could be adopted. I believe this is due to the fact that tactile stimuli are perceived significantly differently by different subjects. Therefore, I believe that the results of the experiments could have been more conclusive if the custom made tactile device was calibrated beforehand for each subject.

Overall this thesis presents innovative evaluation techniques for multimodal environments. Such techniques are inspired by presence research and at the same time they propose new contributions. Therefore, the thesis improves the current state of the art in both understanding what are elements in virtual environments which enhance presence, and in which methods are more suitable to measure presence.

52   CHAPTER 7. CONCLUSIONS

Bibliography

[1] H. Bakstein and T. Pajdla, "Omnidirectional image-based rendering," in *Computer vision winter workshop*. Citeseer, 2006.

[2] D. Rocchesso, S. Serafin, F. Behrendt, N. Bernardini, R. Bresin, G. Eckel, K. Franinovic, T. Hermann, S. Pauletto, P. Susini, et al., "Sonic interaction design: sound, information and experience," in *Proc. CHI*. 2008, ACM New York, NY, USA.

[3] W.A. IJsselsteijn, "Presence in depth," *Eindhoven, The Netherlands: TechnischeUniversiteit Eindhoven*, 2004.

[4] M. Slater, "A note on presence terminology," *Presence connect*, vol. 3, no. 3, 2003.

[5] N.F. Dixon and L. Spitz, "The detection of auditory visual desynchrony," *Perception*, vol. 9, no. 6, pp. 719–721, 1980.

[6] R.L. Storms and M.J. Zyda, "Interactions in perceived quality of auditory-visual displays," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 6, pp. 557–580, 2000.

[7] A. Kohlrausch and S. van de Par, "Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications," in *Proceedings of SPIE*, 1999, vol. 3644, p. 34.

[8] S.S. Stevens and H.E. Pashler, *Steven's Handbook of Experimental Psychology*, John Wiley & Sons, 2002.

[9] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O Society*, vol. 12, no. 7, pp. 35–50, 1992.

[10] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory," *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 282–297, 2001.

54   *BIBLIOGRAPHY*

[11] B.G. Witmer and M.J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[12] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of computer-mediated communication*, vol. 3, no. 2, pp. 20, 1997.

[13] K.M. Stanney, R.R. Mourant, and R.S. Kennedy, "Human factors issues in virtual environments: A review of the literature," *Presence*, vol. 7, no. 4, pp. 327–351, 1998.

[14] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, 1932.

[15] BG Shinn-Cunningham, "Applications of virtual auditory displays," in *Proceedings of the 20th International Conference of the IEEE Engineering in Medicine and Biology Society*, 1998, vol. 20, pp. 1105–1108.

[16] J. Van Baren and W. IJsselsteijn, "Measuring Presence: A Guide to Current Measurement Approaches," *OmniPres project IST-2001-39237, Project Report*, 2001.

[17] J. Freeman, SE Avons, R. Meddis, D.E. Pearson, and W. IJsselsteijn, "Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 2, pp. 149–164, 2000.

[18] M. Slater and A. Steed, "A virtual presence counter," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 5, pp. 413–434, 2000.

[19] A. Steed, A. Brogni, and V. Vinayagamoorthy, "Breaks in presence as usability criteria," in *Proc. of HCI International*. Citeseer, 2005.

[20] R.P. Darken, D. Bernatovich, J.P. Lawson, and B. Peterson, "Quantitative measures of presence in virtual environments: The roles of attention and spatial comprehension," *CyberPsychology & Behavior*, vol. 2, no. 4, pp. 337–347, 1999.

[21] W.A. IJsselsteijn, H. de Ridder, J. Freeman, and S.E. Avons, "Presence: Concept, determinants and measurement," in *PROCEEDINGS-SPIE THE INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING*. Citeseer, 2000, pp. 520–529.

[22] B. ETNYRE and T. KINUGASA, "Postcontraction influences on reaction time," *Research quarterly for exercise and sport*, vol. 73, no. 3, pp. 271–281, 2002.

[23] C.W. Luchies, J. Schiffman, L.G. Richards, M.R. Thompson, D. Bazuin, and A.J. DeYoung, "Effects of age, step direction, and reaction condition on the ability to step quickly," *Journals of Gerontology Series A: Biological and Medical Sciences*, vol. 57, no. 4, pp. 246, 2002.

[24] P. Philip, J. Taillard, P. Sagaspe, C. Valtat, M. Sanchez-Ortuno, N. Moore, A. Charles, and B. Bioulac, "Age, performance and sleep deprivation," *Journal of sleep research*, vol. 13, no. 2, pp. 105–110, 2004.

[25] G. Der and I.J. Deary, "Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey," *Psychology and Aging*, vol. 21, no. 1, pp. 62, 2006.

[26] S. Dane and A. Erzurumluoglu, "Sex and handedness differences in eye-hand visual reaction times in handball players," *International Journal of Neuroscience*, vol. 113, no. 7, pp. 923–929, 2003.

[27] M. Buchsbaum and E. Callaway, "INFLUENCES OF RESPIRATORY CYCLE ON SIMPLE REACTION TIME.," *Perceptual and motor skills*, vol. 20, pp. 961, 1965.

[28] B. Reeves, M. Lombard, and G. Melwani, "Faces on the screen: Pictures or natural experience," in *Annual Conference of the International Communication Association*, 1992.

[29] B. Reeves, B. Detenber, and J. Steuer, "New televisions: The effects of big pictures and big sound on viewer responses to the screen," *Information Systems Division of the International Communication Association, Washington, DC*, 1993.

[30] C. Klimmt, T. Hartmann, A. Gysbers, and P. Vorderer, "The value of reaction-time measures for Presence research: Empirical findings and future perspectives," in *Presence 2005 Conference Proceedings. London, S*, 2005, pp. 293–298.

[31] 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.

[32] J. Pope and A. Chalmers, "Multi-sensory rendering: Combining graphics and acoustics," in *Proceedings of the 7th International Conference in Central Europe on Computer Graphics*. Citeseer, 1999, pp. 233–242.

[33] E.B. Goldstein, B.E. Goldstein, and E.B. Goldstein, *Sensation and perception*, Wadsworth New York, 2002.

[34] J. Freeman and J. Lessiter, "Hear there & everywhere: the effects of multi-channel audio on presence," *Proceedings of ICAD 2001*, pp. 231–234, 2001.

[35] R.D. Sanders Jr, *The effect of sound delivery methods on a userÕs sense of presence in a virtual environment*, Ph.D. thesis, NAVAL POSTGRADUATE SCHOOL, 2002.

[36] C. Porschmann, "Influences of bone conduction and air conduction on the sound of one's own voice," *Acta Acustica united with Acustica*, vol. 86, no. 6, pp. 1038–1045, 2000.

[37] K. Ozawa, S. Ohtake, Y. Suzuki, and T. Sone, "Effects of visual information on auditory presence," *Acoustical Science and Technology*, vol. 24, no. 2, pp. 97–99, 2003.

[38] S. Serafin and G. Serafin, "Sound design to enhance presence in photorealistic virtual reality," in *Proceedings of the 2004 International Conference on Auditory Display*. Citeseer, 2004, pp. 6–9.

[39] P. Chueng and P. Marsden, "Designing auditory spaces to support sense of place: the role of expectation," in *CSCW Workshop: The Role of Place in Shaping Virtual Community*. Citeseer, 2002.

[40] C. Porschmann, "One's own voice in auditory virtual environments," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 378–388, 2001.

[41] K. Ozawa and M. Miyasaka, "Effects of reproduced sound pressure levels on auditory presence," *Acoustical Science and Technology*, vol. 25, no. 3, pp. 207–209, 2004.

[42] C.D. Murray, P. Arnold, and B. Thornton, "Presence accompanying induced hearing loss: Implications for immersive virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 2, pp. 137–148, 2000.

[43] 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.

[44] P. Larsson, "Virtually hearing, seeing, and being: room acoustics, presence, and audiovisual environments," *DOKTORSAVHANDLINGAR-CHALMERS TEKNISKA HOGSKOLA*, vol. 2319, 2005.

[45] J.M. Knapp and J.M. Loomis, "Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 13, no. 5, pp. 572–577, 2004.

[46] P. Larsson, D. Vastfjall, and M. Kleiner, "Perception of self-motion and presence in auditory virtual environments," in *Proceedings of seventh annual workshop presence*, 2004, pp. 252–258.

[47] B. Kapralos, D. Zikovitz, M. Jenkin, and L.R. Harris, "Auditory cues in the perception of self-motion," in *116th AES convention*. Citeseer, 2004.

[48] A. Valjamae, P. Larsson, D. Vastfjall, and M. Kleiner, "Travelling without moving: Auditory scene cues for translational self-motion," in *Proceedings of ICADÕ05*, 2005.

[49] M. Chion, C. Gorbman, and W. Murch, *Audio-vision: sound on screen*, Columbia Univ Pr, 1994.

BIBLIOGRAPHY 57

[50] M. Leman, *Embodied music cognition and mediation technology*, The MIT Press, 2007.

[51] R.I. Godoy, *Musical Gestures: Sound, Movement, and Meaning*, Routledge, 2009.

[52] P. Larsson, D. V
"astfj
"all, and M. Kleiner, "The actor-observer effect in virtual reality presentations," *CyberPsychology & Behavior*, vol. 4, no. 2, pp. 239–246, 2001.

[53] FW Galbraith and MV Barton, "Ground loading from footsteps," *The Journal of the Acoustical Society of America*, vol. 48, pp. 1288, 1970.

[54] A. Ekimov and J.M. Sabatier, "Vibration and sound signatures of human footsteps in buildings," *The Journal of the Acoustical Society of America*, vol. 120, pp. 762, 2006.

[55] X. Li, R.J. Logan, and R.E. Pastore, "Perception of acoustic source characteristics: Walking sounds," *Journal of the Acoustical Society of America*, vol. 90, no. 6, pp. 3036–3049, 1991.

[56] JM Sabatier and AE Ekimov, "A Review of Human Signatures in Urban Environments Using Seismic and Acoustic Methods," in *Technologies for Homeland Security, 2008 IEEE Conference on*, 2008, pp. 215–220.

[57] G. Wu and J.H. Chiang, "The effects of surface compliance on foot pressure in stance," *Gait & Posture*, vol. 4, no. 2, pp. 122–129, 1996.
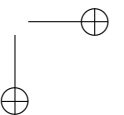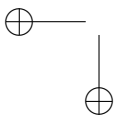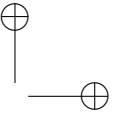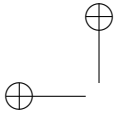
58   *BIBLIOGRAPHY*

Part II

PAPERS

PAPER I

R. Nordahl

Evaluating environmental sounds from a presence
perspective for virtual reality applications

Evaluating environmental sounds from a presence perspective for virtual reality applications

Rolf Nordahl

Medialogy, Aalborg University Copenhagen,
Lautrupvang 15, 2750 Ballerup, DK
rn@media.aau.dk

ABSTRACT

In this paper, 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 bi-modal 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 ego-motion 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 [23]. While research has provided different results concerning multimodal interaction among the senses [24], 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 uni-modality (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 [14].

Engineers have been interested in the audio-visual interaction from the perspective of optimizing the perception of quality offered by technologies [8, 20]. Furthermore, studies have shown that by utilizing audio, the perceived quality of lower quality visual displays can increase [25]. Likewise, researchers from neuroscience and psychology have been interested in the multimodal perception of the auditory and visual senses [13]. 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 [2] and the ventriloquism effect [10].

The design of immersive virtual environments is a challenging task, and cross-modal stimulation is an important tool for achieving this goal [16]. However, the visual modality is still dominant in VR technologies. A common approach when designing multimodal systems consists of adding other sensorial stimulation 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 [22]. 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, 3]. 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 [3].

In [14], 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 en-

hancing sense of presence has been outlined by several researchers [11, 22, 9]. Most of the research relating sound and presence has examined the role of sound versus non sound and the importance of spatial qualities of the auditory feedback.

In [19], 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, i.e., footsteps, vehicles, doors etc. 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 [5, 21]. Taking the approach of ecological perception, in [5] 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 [18]. The goal was to advocate the use of interactive auditory feedback as a mean to enhance motion of subjects and sense of presence in a photorealistic virtual environment. We focused both on ambient sounds, defined as sounds characteristic 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, i.e., 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") [1], completed in 2005, had as its main focus the development of new synthetic image rendering technologies (commonly referred to as Image Base Rendering (IBR)) that allowed photo-realistic 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 multi-level 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, i.e., 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 360^0 inside the so-called region of exploration (REX). However, no events happen in the environment, which make it rather uninteresting to explore.

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

---

[1]www.benogo.dk

walk. This is an important observation for when sonically simulating the act of walking in the environment.

Figure 1: An image of the Prague botanical garden used as visual feedback in the experiments.

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 [1].

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) [6].

The footstep synthesizer was built starting by analyzing footsteps recorded on surfaces obtained from the Hollywood Edge Sound Effects library.[2]. For each recorded sets of sounds, single steps were isolated and analyzed. The main goal of the analysis was to identify an average amplitude

---
[2]www.hollywoodedge.com

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.[3] The sensors were wirelessly connected to a microcontroller, as shown in Figure 2, and the microprocessor was connected to a laptop PC.

The continuos pressure value was used to control the force of the impact of each foot on the floor, to vary the temporal evolution of the synthetic 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 creating an 8-channels 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[7], where eight channels surround is presently implemented, in order to perform in the future experiments with higher quality visual feedback. This is the reason why 8-channels sound rendering was chosen compared to e.g., binaural rendering [4].

Three kinds of auditory feedback were implemented:

1. "Static" soundscape, reproduced at max. peak of 58dB, measured c-weighted with slow response. This soundscape was delivered through the 8-channels system.
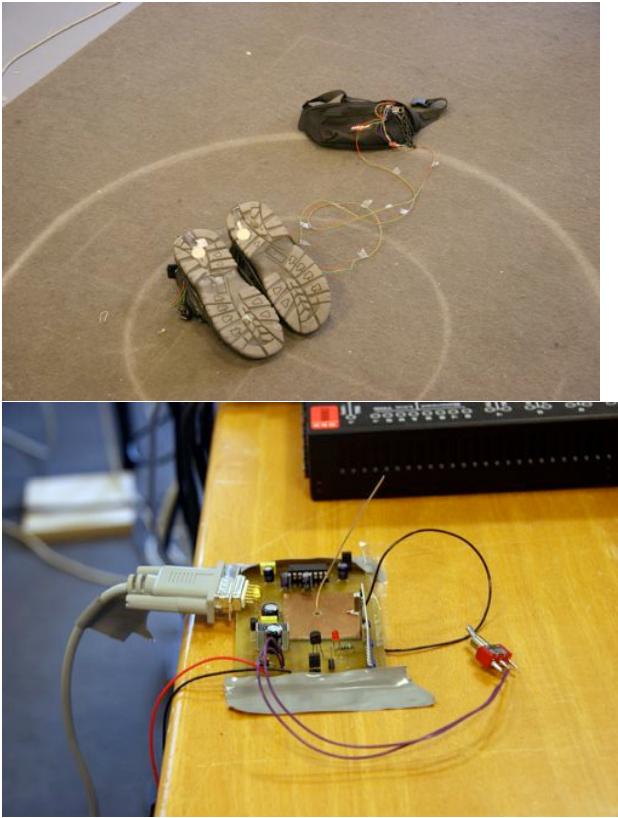
---
[3]www.cycling74.com

Figure 2: The sandals (top) enhanced with pressure sensitive sensors wirelessly connected to a microprocessor (bottom).

Figure 3: A subject navigating in the virtual environment wearing an head mounted display (HMD).

2. Dynamic soundscape with moving sound sources, developed using the VBAP algorithm, 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 [17])

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, 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, infact, the position of the sound sources was dynamic and controlled by the user's motion, who was wearing an head tracker as described below. In the third condition, the dynamic 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
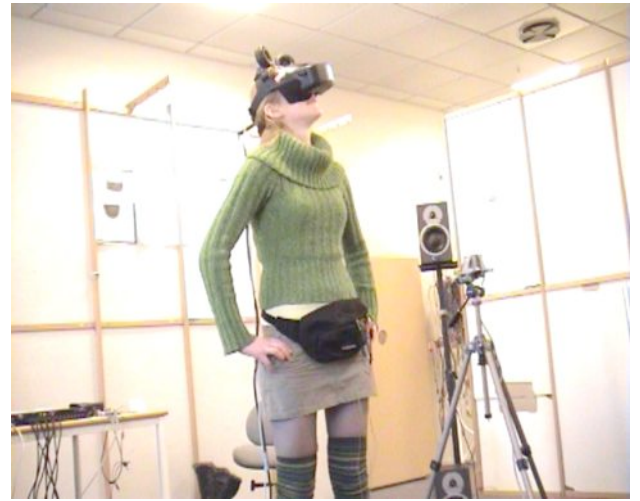
visual feedback, and another one the auditory feedback together with the interactive shoes. A Polhemus tracker (Iso-Track II3), attached to the head mounted display, was connected to the computer running the visual display, and allowed to track the position and orientation of the user in 3D. The computer 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 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: ((640x3)x480), (921,600 color elements) equivalent to 307,200 triads. Furthermore the HMD provides a field of view of 60^0 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[4] was used. Sound was delivered by eight Dynaudio BM5A speakers.[5] 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.

---

[4]http://www.rmeaudio.com/english/firewire/
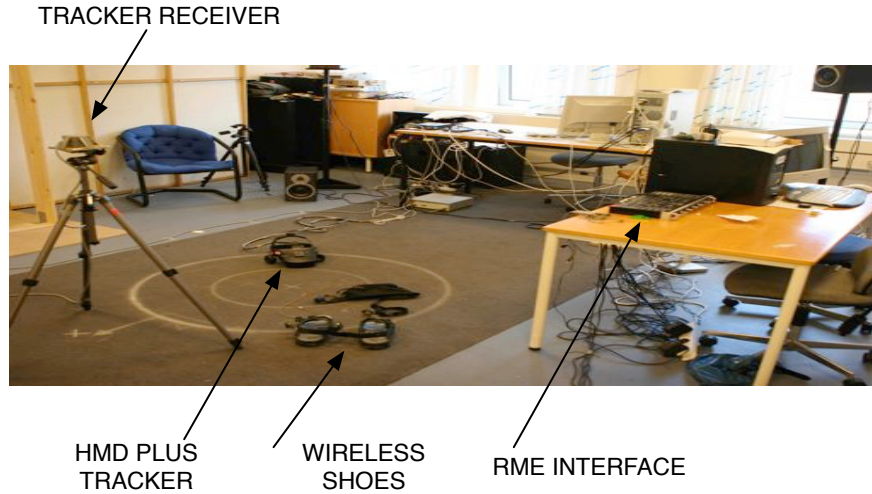[5]http://www.dynaudioacoustics.com

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.
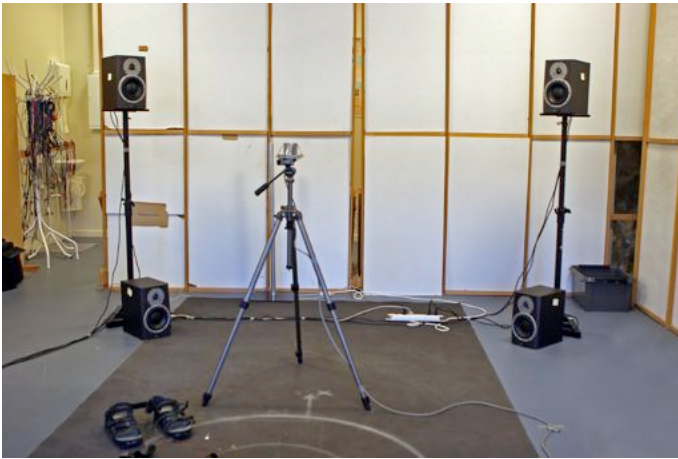

Figure 5: A different view of the 8-channels surround sound lab where the experiments were run.

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 to 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 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.

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 [17].

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 the VBAP algorithm) 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.

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 music[6] described before.

---

[6]Mozart, Wolfgang Amadeus, Piano Quintet in E flat, K. 452, 1. Largo Allegro Moderato, Philips Digitals Classics, 446 236-2, 1987
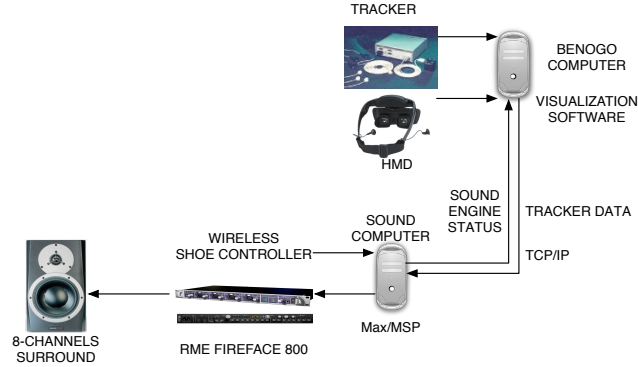
Figure 6: Connection of the different hardware components in the experimental setup.

CONDITION	AUDITORY STIMULI	NUM 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 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.

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

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

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.

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.

7. RESULTS

Table 7 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 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 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

	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	0.41	0.197	0.003	0.018	0.347	

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

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

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

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

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 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 is introduced. (*Full* compared to *Visual Only* (p=0.005), *Full seq* compared to *Visual Only* (p=0.051)).

Figure 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 (top plot in Figure 7) is clearly noticeable. The subject in the full condition (bottom plot in Figure 7) 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 (center plot in Figure 7).

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 [15], from 1-7, (from 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 feed-

back 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 significance difference with the second highest ranked condition in the list (full seq., mean=71.4) ($p < 0.02$). 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, i.e., 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. 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 sound 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 that it is the combination of soundscapes, 3-dimensional sound and auditory rendering of ones own motion that induces a higher degree of motion.

We also investigated if 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.
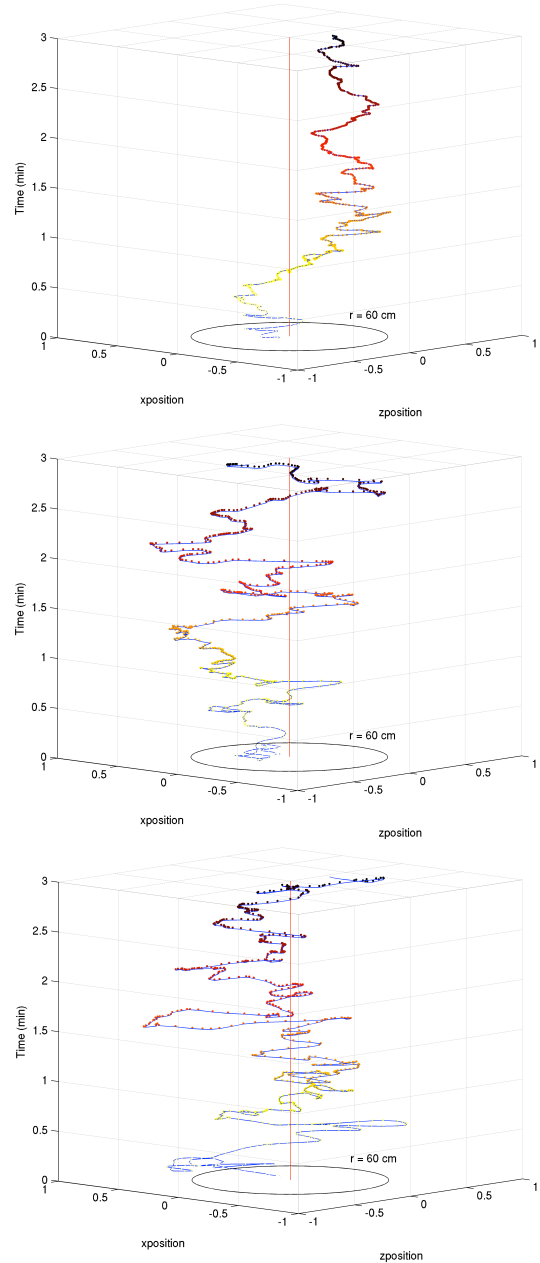

Figure 7: Visualization over time of the motion of one subject in the six different conditions. From top to bottom: visual, visual w. foot, full.

10. REFERENCES

[1] J.M. Adrien. The missing link: modal synthesis. *Representations of musical signals table of contents*, pages 269–298, 1991.

[2] B. Arons. A review of the cocktail party effect. *Journal of the American Voice*, 12, 1992.

[3] Witmer B. and Singer M. Measuring Presence in Virtual Environments. A Presence Questionnaire. *Presence*, 7(3):225–240, 1998.

[4] D.R. Begault. *3-D sound for virtual reality and multimedia.* AP Professional Boston, 1994.

[5] P. Chueng and P. Marsden. Designing auditory spaces to support sense of place: the role of expectation. In *CSCW Workshop: The Role of Place in Shaping Virtual Community.* Citeseer, 2002.

[6] P.R. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

[7] C. Cruz-Neira, D.J. Sandin, T.A. DeFanti, R.V. Kenyon, and J.C. Hart. The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992.

[8] N.F. Dixon and L. Spitz. The detection of audiovisual desynchrony. *Perception*, 9, 1980.

[9] 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*, 4(4):357–363, 1995.

[10] S. Handel. *Perceptual coherence: hearing and seeing.* Oxford University Press, USA, 2006.

[11] W.A. IJsselsteijn. Presence in depth. *Eindhoven, The Netherlands: TechnischeUniversiteit Eindhoven*, 2004.

[12] Lessiter J. A Cross Media Questionnaire - The ITC Sense of Presence Inventory. *Presence*, 10(3):282–297, 2001.

[13] A Kohlrausch and S. Vand de. Par. Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications. In *Proceeding of IST/SPIE Conference on Human Vision and Electronic Imaging IV*, 1999.

[14] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):20, 1997.

[15] T.J. Maurer and H.R. Pierce. A comparison of Likert scale and traditional measures of self-efficacy. *Journal of applied psychology*, 83(2):324–329, 1998.

[16] Durlach N. and Mavor A. *Virtual reality: Scientific and Technological challenges.* National Academy Press, 1995.

[17] R. Nordahl. Auditory rendering of self-induced motion in virtual reality. *M. Sc. project report, Dept. of Medialogy, Aalborg University Copenhagen*, 2005.

[18] R. Nordahl. Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion. *Proceedings of Presence*, pages 57–62, 2006.

[19] K. Ozawa, Y. Chujo, Y. Suzuki, and T. Sone. Psychological factors involved in auditory presence. *Acoustical Science and Technology*, 24(1):42–44, 2003.

[20] 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.

[21] S. Serafin and G. Serafin. Sound design to enhance presence in photorealistic virtual reality. In *Proceedings of the 2004 International Conference on Auditory Display*, pages 6–9. Citeseer, 2004.

[22] M. Slater. A note on presence terminology. *Presence connect*, 3(3), 2003.

[23] K.M. Stanney et al. *Handbook of virtual environments: design, implementation, and applications.* Lawrence Erlbaum Associates, 2002.

[24] B.E. Stein, M.A. Meredith, and S. Wolf. The merging of the senses. 1993.

[25] R.L. Storms and M.J. Zyda. Interactions in perceived quality of auditory visual displays. *Presence*, 9, 2000.

[26] D. Västfjäll, P. Larsson, and M. Kleiner. Development and validation of the Swedish viewer-user presence questionnaire (SVUP), 2000.
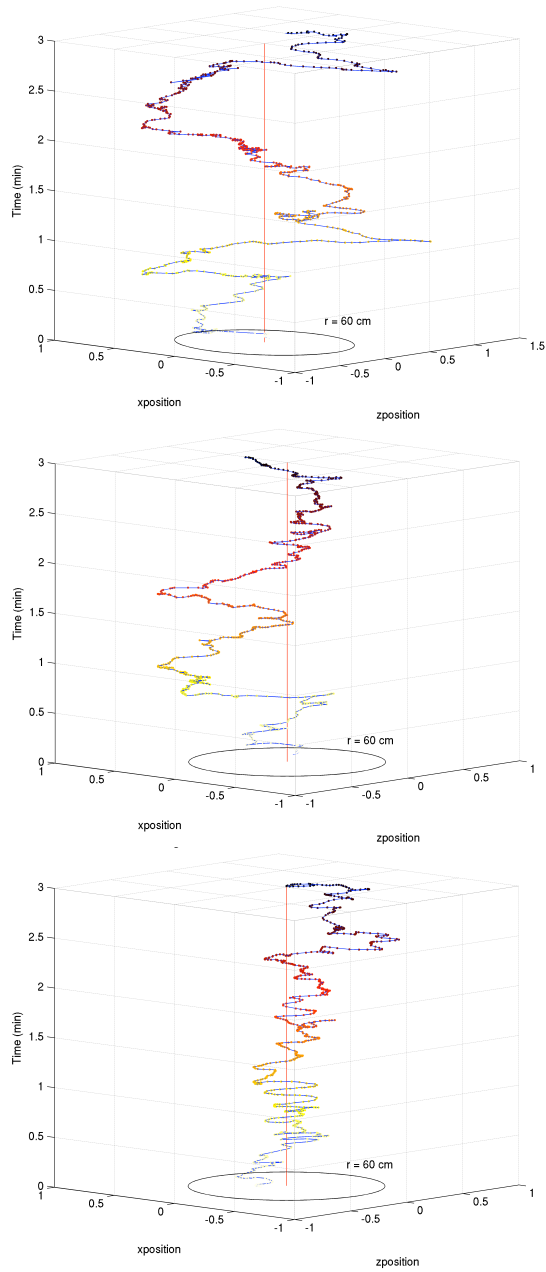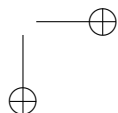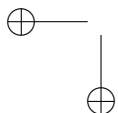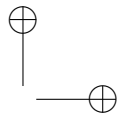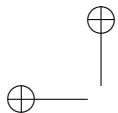
Figure 8: Visualization over time of the motion of one subject in the six different conditions. From top to bottom: full sequenced, sound+3D and music.

PAPER II

R. Nordahl, S. Serafin and L. Turchet.

Sound synthesis and evaluation of interactive footsteps for virtual reality applications

Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications

Rolf Nordahl *   Stefania Serafin†   Luca Turchet‡

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

ABSTRACT

A system to synthesize in real-time the sound of footsteps on different materials is presented. The system is based on microphones which allow the user to interact with his own footwear. This solution distinguishes our system from previous efforts that require specific shoes enhanced with sensors. The microphones detect real footsteps sounds from users, from which the ground reaction force (GRF) is estimated. Such GRF is used to control a sound synthesis engine based on physical models. Evaluations of the system in terms of sound validity and fidelity of interaction are described.

Keywords: sound synthesis, physical models, footsteps sounds, auditory perception

Index Terms: H.5.5 [Information Systems]: Information Interfaces and Presentation—Sound and Music Computing; H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

1 INTRODUCTION

The development of efficient yet accurate simulation algorithms, together with improvements in hardware technology, has boosted the research on auditory display and physically based sound models for virtual environments (VEs)[27, 23, 7].

The addition of auditory cues and their importance in enhancing the sense of immersion and presence is a recognized fact in virtual environment research and development. Most prior work in this area has focused on sound delivery methods [25, 24], sound quantity and quality of auditory versus visual information [4] and 3D sound [11, 28]. Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in virtual environments [17, 15, 26].

Self-generated sounds have been often used as enhancements to VEs and first-person 3D computer games – particularly in the form of footstep sounds accompanying self-motion or the presence of other virtual humans.

Such sounds are used to produce embodiment and a sense of weight with the overall goal of heightening the sense of "realness" to the character or person. Usually such sounds are taken from sound libraries or recorded by *Foley* artists who put shoes in their hands and interact with different materials to simulate the act of walking.

Recently, several physics based algorithms have been proposed to simulate the sounds of walking. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [5]. Among such algorithms the sounds of people

---

*e-mail: rn@media.aau.dk
†e-mail:sts@media.aau.dk
‡e-mail:tur@media.aau.dk

walking on different surfaces were simulated [6]. A similar algorithm was also proposed in [10], where physically informed models reproduced several aggregate surfaces. Procedural sound synthesis of walking has also been recently described in [9].

Previous work on interactive footwear, such as the research performed by Paradiso and coworkers [21, 3], consisted of designing shoes augmented with sensors used to control footsteps sounds. A smaller number of examples, such as recent work of Nordahl [20] and Law et al. [18], have even aimed to provide multimodal cues linked to footsteps events in such environments.

In this paper, we are particularly interested in developing a solution which requires a minimum amount of sensing technology and is shoe independent, which means that subjects can keep their own footwear while using the system. This creates several advantages from the interaction side: users do not need to wear ad-hoc designed shoes, whose wearability is decreased by the addition of several sensors.

We propose an interactive system which enables a designer to synthesize in real-time footsteps sounds of different materials. We describe the results of experiments whose goal is to test the degree of realism of the system, the ability of subjects to recognize the virtual material they are walking on, and the fidelity of interaction. The ultimate goal is to integrate this system in the simulation of multimodal virtual environments where the act of walking plays an important role.

2 THE SOUND SYNTHESIS ENGINE

We developed a physically based sound synthesis engine able to simulate the sounds of walking on different surfaces. Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades [12]. In our simulations, we draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

2.1 Solid surfaces

Sonic interactions between solid surfaces have been extensively investigated, and results are available which describe the relationship between physical and perceptual parameters of objects in contact [16, 27]. Such sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay.

A common approach to synthesizing such sounds is based on a lumped source-filter model, in which an impulsive excitation $s(t)$, modeling the physics of contact, is passed through a linear filter $h(t)$, modeling the response of the vibrating object as $y(t) = s(t) \star h(t)$.

Modal synthesis [1] is one widely adopted implementation of this idea. In this synthesis technique, the response model $h(t)$ is decomposed in terms of the resonant frequencies f_i of the vibrating object, also known as the modes of the object. The re-

sponse is modeled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, where a_i represent the amplitudes of the modes, b_i the decay rates of the modes, and f_i the frequencies of the modes.

A footstep sound can be considered as the result of multiple micro-impact sounds between a shoe and a floor. The set of such micro-events can be thought as the result of the interaction between an *exciter* and a *resonator*. The exciter is represented by the interaction between shoe and ground. Such interaction can be either continuous, as in the case of a foot sliding across the floor, or discrete, as in the case of walking on a solid surface.

To simulate such scenarios, both an impact and friction model were implemented.

In the impact model, the excitation corresponding to each impact $s(t)$ is assumed to possess a short temporal extent and an unbiased frequency response. Such excitation consists of a discrete-time model of the force f between the two bodies, dependent on additional parameters governing the elasticity of the materials, their velocity of impact $\dot{x}$, and masses:

$$f(x,\dot{x}) = \begin{cases} -kx^{\alpha} - \lambda\, x^{\alpha}\, \dot{x} & \text{if } x > 0 \\ 0 & x \leq 0 \end{cases}$$

where α depends on the local geometry around the contact surface, and x stands for the compression of the exciter (when $x > 0$ the two objects are in contact) [2].

In the friction model we adopted a dynamic model, where the relationship between relative velocity v of the bodies in contact and friction force f is represented through a differential equation rather than static mapping. Assuming that friction results from a large number of microscopic elastic bonds, called bristles in [8], the v-to-f relationship is expressed as:

$$f(z,\dot{z},v,w) = \sigma_0 z + \sigma_1 \dot{z} + \sigma_2 v + \sigma_3 w$$

where z is the average bristle deflection, the coefficient σ_0 is the bristle stiffness, σ_1 the bristle damping, and the term $\sigma_2 v$ accounts for linear viscous friction. The fourth component $\sigma_3 w$ relates to surface roughness, and is simulated as fractal noise.

2.2 Aggregate surfaces

To synthesize aggregate surfaces, we implemented the physically informed sonic models (PhiSM) algorithm [5].

This model simulates particle interactions by using a stochastic parameterization. This means that the different particles do not have to be modeled explicitly, but only the probability that particles will create some noise is simulated. For many particle systems, this phenomenon is well taken into account by using a simple Poisson distribution, where the sound probability is constant at each time step, giving rise to an exponential probability waiting time between events.

2.3 Implementation

Using the algorithms described in the previous sections, we implemented a comprehensive collection of footstep sounds. As solid surfaces, we implemented metal and wood. In these materials, the impact model was used to simulate the act of walking, while the friction model was used to simulate the sound of creaking wood.

As aggregate surfaces, we implemented gravel, sand, snow, forest underbrush, dry leaves, pebbles and high grass. The simulated metal, wood and creaking wood surfaces were furthermore enhanced by using some reverberation. The role of reverberation is discussed in the testing section.

The sound synthesis algorithms were implemented in C++ as external libraries for the Max/MSP sound synthesis and multimedia

real-time platform.[1] To enable compatibility with the Pure Data platform,[2] the algorithms were implemented using Flext.[3]

In our simulations, designers have access to a sonic palette making it possible to manipulate all such parameters, including material properties. One of the challenges in implementing the sounds of different surfaces was to find suitable combinations of parameters which provided a realistic simulation. For each simulated surface, recorded sounds were analyzed according to their combinations of events, and each subevent was simulated independently. As an example, the sound produced while walking on dry leaves is a combination of granular sounds with long duration both at low and high frequencies, and noticeable random sounds with not very high density that give to the whole sound a crunchy aspect. These different components were simulated with several stochastic models having the same density, duration, frequency and number of colliding objects.

The amplitude of the different components were also appropriately weighed, according to the same contribution present in the corresponding real sounds. Finally, a scaling factor for the subcomponents volumes gives to the whole sound an appropriate volume, in order to recreate a similar sound level which it would happen during a real footstep on each particular material.

A pilot test was run to ascertain that such a global volume plays an important role in the judgments concerning the sounds' realism and in the recognition of the surface material. Indeed, wrong settings for such a parameter can cause wrong recognitions.

3 CONTROLLING THE SOUND SYNTHESIS ENGINE

The developed sound synthesis engine is controlled as following: users are asked to walk inside an area delimited by four microphones placed on the floor in a square configuration. Specifically, we used four Shure BETA 91,[4] high performance condenser microphones with a tailored frequency response designed specifically for kick drums and other bass instruments. The microphones' features made them a good candidate for the purpose of capturing footsteps sounds. In the interaction between a foot and a sole, the exciter is usually called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. The aim of the phase of analysis has been that of extracting the GRF from the acoustic waveform. The real footsteps sounds produced are detected by the microphone, and their GRF extracted and used to control the temporal evolution of the synthetic footsteps. An example of a footstep sound and its corresponding GRF is shown in Figure 1.

4 EXPERIMENT

We conducted different experiments whose goal is to investigate the ability of subjects to recognize the different walking sounds they were exposed to. The study of human perception of locomotion sounds has addressed several properties of walking sound sources: the gender [19, 13] and posture of a walker [22], the emotions of a walker [13], the hardness and size of the shoe sole [13], and the ground material [14].

Such studies have been concerned only with recognition of sounds in an off-line scenario, where subjects were asked to listen to some sounds and classify them. In this experiment, we are interested in having subjects classify sounds both off-line, but also in an active settings, i.e., by using the developed interactive system. One of our hypotheses is that the recognition when using the interactive system is higher than in the off-line setup.

Moreover, we conducted an experiment using recorded real sounds in order to compare their recognition rate with that of the developed synthesized sounds.

---

[1] www.cycling74.com

[2] www.puredata.org

[3] http://puredata.info/Members/thomas/flext
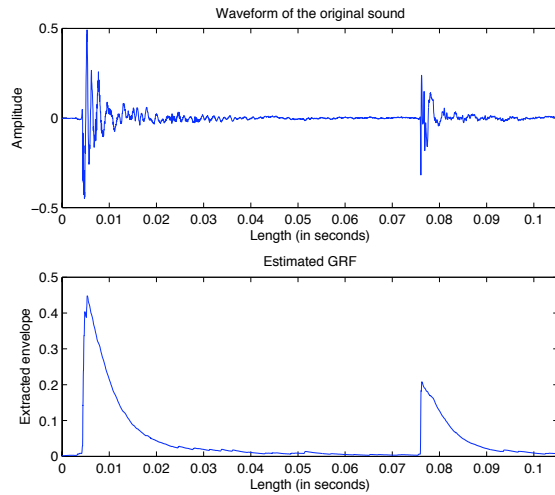
[4] http://www.shure.com/

Figure 1: A footstep sound (top) and the corresponding calculated GRF (bottom).

4.1 Methods

Three kinds of experiments were conducted:

1. experiment 1: recognition of footsteps sounds generated in real time by the subjects.

2. Experiment 2: recognition of synthesized recorded footsteps sounds.

3. Experiment 3: recognition of real recorded footsteps sounds.

The sounds provided during experiment 1 were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section.
The sounds provided during experiment 2 consisted of recordings of footsteps sounds generated by the use of the interactive system.
The sounds provided during experiment 3 consisted of recordings of real footsteps sounds on different surfaces. Such sounds were chosen among those available on the Hollywood Edge sound effects library.[5] Each sound in experiment 2 and 3, composed of several footsteps, had duration of about 7 seconds.

Participants were exposed to 26 trials in experiment 1 and 2, and 30 trials in experiment 3. During experiments 1 and 2, 13 stimuli were presented twice in randomized order. The stimuli consisted of footsteps sounds on the following surfaces: beach sand, gravel, dirt plus pebbles (like in a country road), snow (in particular deep snow), high grass, forest underbrush (a forest floor composed by dirt, leaves and branches breaking), dry leaves, wood, creaking wood and metal. To simulate room characteristics, footsteps sounds on wood, creaking wood and metal were enhanced adding a certain amount of reverberation.

In experiment 3, fifteen stimuli were presented twice in randomized order. They consisted of the previous mentioned sounds without the reverberated ones, more footsteps sounds on carpet, concrete, frozen snow, puddles and water.

4.1.1 Participants

Fourtyfive participants were divided in three groups (n = 15) to perform the three between-subjects experiments. The three groups

[5] www.hollywoodedge.com/

were composed respectively of 6 men and 9 women, aged between 19 and 29 (mean=22.13,standard deviation=2.47), 8 men and 7 women, aged between 20 and 35 (mean=22.73,standard deviation=4.01), and 10 men and 5 women, aged between 20 and 29 (mean=23.13,standard deviation=2.39). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

During experiment 1 the shoes used by subjects were sneakers, trainers, boots and other kinds of shoes with rubber soil.
The participants took in average about 24, 15 and 12 minutes for experiments 1, 2 and 3 respectively.
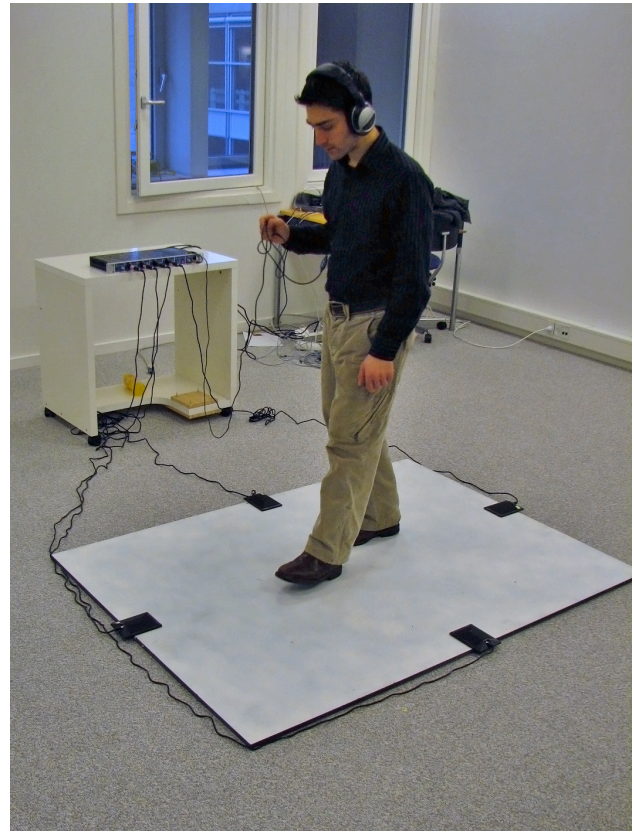
4.1.2 Setup

Figure 2: A subject performing experiment 1.

All experiments were carried out in an acoustically isolated laboratory where three setups were installed. The setup for experiment 1 consisted of the interactive system, and the participants were asked to use it in order to generate the footsteps sounds in real time (see Figure 2).
The setup for experiments 2 and 3 consisted of a simple graphical user interface with which the participants were asked to interact, and a spreadsheet to collect their answers.

The interface was created using the Max/MSP program and was composed only by buttons to be pressed. Each button was numbered, and by pressing it a sound was triggered and conveyed to the user by means of headphones. Users were asked to press each button according to their numerical order, and to write the corresponding answers on the spreadsheet.

	Correct answers	Wrong answers	I don't know	Realism	Quality
Beach Sand	50.	36.67	13.33	5.3	5.15
Gravel	70.	26.66	3.33	5.57	5.57
Dirt pebbles	10.	86.67	3.33	5.75	4.73
Snow	80.	16.67	3.33	5.81	5.67
High Grass	0.	83.33	16.67	-	5.4
Forest Underbrush	63.33	33.33	3.33	4.84	4.7
Dry Leaves	40.	60.	0.	4.75	4.96
Wood	46.67	20.	33.33	3.93	4.55
Creaking Wood	93.33	6.67	0.	5.16	5.17
Metal	80.	13.33	6.67	3.33	4.77
Wood plus Reverb	20.	70.	10.	3.83	4.76
Creaking Wood plus Reverb	93.33	3.33	3.33	4.93	5.17
Metal plus Reverb	83.33	10.	6.67	3.6	5.23

Table 1: Results of experiment 1: recognition (in percentage) of the surfaces with the interactive system.

4.1.3 Task

During experiment 1 the participants were asked to wear a pair of headphones and to walk in the area delimited by the microphones. They were given the list of different surfaces to be held in one hand, presented as non-forced alternate choice.

During the act of walking they listened simultaneously to footsteps sounds on a different surface according to the stimulus presented. The task consisted of answering by voice the following three questions after the presentation of the stimulus:

1. Which surface do you think you are walking on? For each stimulus choose an answer in the following list: 1) beach sand, 2) gravel, 3) dirt plus pebbles, 4) snow, 5) high grass, 6) forest underbrush, 7) dry leaves, 8) wood, 9) creaking wood, 10) metal, 11) carpet, 12) concrete, 13) frozen snow, 14) puddles and water, 15) I don't know.

2. How close to real life is the sound in comparison with the surface you think it is? Evaluate the degree of realism on a scale from 1 to 7 (1=low realism, 7=high realism).

3. Evaluate the quality of the sound on a scale from 1 to 7 (1=low quality, 7=high quality).

At the end of the experiment, subjects were asked some questions concerning the naturalness of the interaction with the system and to comment on its usability and possible integration in a virtual reality environment.

The task in experiment 2 was similar to experiment 1. However, subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface mentioned in section 4.1.2. The task in experiment 3 was similar to experiment 2, but in addition to the classification of the surfaces subjects were also asked to evaluate the degree of certainty of their choice on a scale from 1 to 7. At the end of the experiments 2 and 3 the subjects were also given the opportunity to leave an open comment on their experience interacting with the system.
The list included a range of materials wider than those presented in experiment 1 and 2 (see section 4.1). Conversely, in experiment 3 all the materials in the list were presented. The subjects were informed that they could choose the same material more than one time and that they were not forced to choose all the materials in the list. In addition for experiment 1, they could use the interactive system as much as they wanted before giving an answer. Likewise for experiments 2 and 3 they could listen to the sounds as much as they wanted. When passed to the next stimulus they could not change the answer to the previous stimuli.

4.2 Results

The collected answers were analyzed and compared between the three experiments. Results concerning the percentage of correct answers in experiment 1 and 2 are illustrated in Figure 3, while the comparison between the two experiments in terms of realism and quality of the sounds is showed in Figure 4. The degree of realism was calculated only looking at data from correct answers, i.e., when the surface was correctly recognised.

The first noticeable element emerging from both figures is that almost always the use of the interactive system gave rise to a better recognition of the surfaces and a higher evaluation of realism and quality of the proposed sounds, rather than the recorded sounds.
In both experiments the footsteps sounds on snow, creaking wood (with and without reverb), gravel and metal (with reverberation) were correctly recognized with high percentage, while the recognition of the surfaces dirt plus pebbles, high grass and wood (with reverberation) turned out to be wrong most of the times. Regarding the recognition of the other surfaces, good results were found for beach sand and forest underbrush, while correct recognition for dry leaves and wood (without reverberation) were under 50%.

All percentages were higher in experiment 1, although an in-depth analysis shows significant difference only for dry leaves (χ^2 = 4.1761, df = 1, p-value = 0.041) and metal (χ^2 = 4.6886, df = 1, p-value = 0.03036).

An analysis performed on the wrong answers reveals that in average subjects tended to classify erroneously a surface as another belonging to a same category (e.g., wood-concrete, snow-frozen snow, dry leaves-forest underbrush) rather than to different categories (e.g., wood-water, wood-gravel, metal-dry leaves).
Moreover, results show that the addition of the reverberation to the sounds gave rise to better recognitions for metal, and worse for wood plus reverberation, which was perceived most of the times as concrete (not tangible differences were found for the creaking wood). As concerns the comparisons between reverberated and not reverberated sounds in terms of realism negligible differences were found, while in terms of quality the reverberated sounds led to light higher evaluations.

Results of the third experiment are illustrated in Figure 5. Recognition of recorded sounds was quite good in average, with a better performance for the solids and liquids surfaces compared to the aggregate ones. In particular metal, wood, creaking wood, concrete, frozen snow, water and gravel show very high percentages, while the sound of the high grass confirms the negative trend already emerged in the previous experiments. All the other materials present percentages over 50%, with the exception of dry leaves, dirt plus pebbles and forest underbrush, as the data in Table 3 show.

	Correct answers	Wrong answers	I don't know	Realism	Quality
Beach Sand	36.67	63.33	0.	4.18	4.77
Gravel	66.67	33.33	0.	4.9	4.57
Dirt pebbles	10.	90.	0.	4.71	4.37
Snow	83.33	13.33	3.33	5.17	5.14
High Grass	3.33	93.33	3.33	5.	4.3
Forest Underbrush	36.67	60.	3.33	4.22	4.27
Dry Leaves	13.33	86.67	0.	4.5	3.87
Wood	36.67	33.33	30.	4.58	4.
Creaking Wood	76.67	13.33	10.	3.17	3.82
Metal	50.	16.67	33.33	2.93	3.29
Wood plus Reverb	6.67	73.33	20.	6.	4.04
Creaking Wood plus Reverb	76.67	13.33	10.	3.35	3.89
Metal plus Reverb	66.67	6.67	26.67	3.3	4.3

Table 2: Results of experiment 2: recognition (in percentage) of the surfaces with the recorded synthesized sounds.
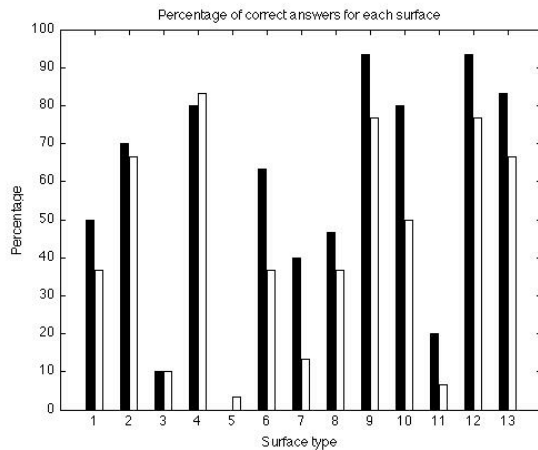

Figure 3: Comparison of the percentages of correct answers for each surface in experiment 1 (black) and 2 (white). Surface type from left to right: 1-beach sand, 2-gravel, 3-dirt pebbles, 4-snow, 5-high grass, 6-forest underbrush, 7-dry leaves, 8-wood, 9-creaking wood, 10-metal, 11-wood plus reverberation, 12-creaking wood plus reverberation, 13-metal plus reverberation. Notice that the missing element in column 5 indicates the fact that none of the subjects was able to recognize the high-grass in experiment 1.

The degree of certainty in the answers seems to be on average consistent with the percentage of correctness (even if there are some exceptions, as for the footsteps on puddles, which were erroneously classified as footsteps through the water).

What emerges from these results is the ability of the subjects in distinguishing materials in the same category for solid surfaces, and their difficulties in the recognition of aggregate surfaces (aspect also confirmed by the comments of the participants). Indeed the analysis of the wrong answers for aggregate surfaces confirms the tendency already showed in the previous experiments, in classifying erroneously a surface as another belonging to a same category.

From the comparison with the results of the recognition of the surfaces presented in the previous two experiments, one can note that the percentage of correct answers for the same surfaces is higher for the experiment 3, with the exception of the snow in both experiment 1 and 2, and of forest underbrush and creaking wood (with reverberation) in experiment 1. Moreover similar percentages were found for beach sand in experiment 1 and 3, and the percentage of gravel is high in the same experiments. Finally the very low percentages for the high grass in the three experiments confirm that this is a sound difficult to recognize.

The final questions of experiment 1 (evaluated on a seven-point Likert scale) show that subjects judged the interaction with the system quite natural (mean= 5.6), and that they felt quite normal (mean= 5.33) and a little bit constrained (mean 2.9) during the act of walking. Indeed, subjects commented on the need of a wider area to walk and of a wireless headphones set.

Finally, regarding the "I do not know" answers the percentage was higher in experiment 2 (10.77%) rather than experiment 1 (7.95%), and lower (3.11%) for the experiment 3. Tables 1,2, 3 show in details the results of experiment 1, 2 and 3 respectively.

5 DISCUSSIONS

A footstep sound is extremely dependent both on the kind of shoes a person is wearing and on the kind of floor the person is walking on. All sounds were synthesized assuming that the shoes hitting the floor had a solid sole. This aspect is extremely important in the simulation of solid floors. As a matter of fact, when interacting with virtual wood and metal more than one participant commented of having the sensation of wearing a different kind of shoe. More precisely, they commented that they felt like they were wearing a shoe with a solid sole. This indicates the ability of auditory feedback to affect perception of material.

homogeneous ones: as proof of the good success our design we

	Correct answers	Wrong answers	I don't know	Degree of certainty
Beach Sand	53.34	46.67	0.	4.5
Gravel	86.67	13.33	0.	5.07
Dirt pebbles	43.33	56.67	0.	4.8
Snow	63.33	36.67	0.	5.93
Frozen Snow	70.	16.67	13.33	5.66
High Grass	10.	83.33	6.67	3.5
Forest Underbrush	36.67	60.	3.33	4.14
Dry Leaves	56.67	43.33	0.	4.77
Concrete	70.	23.33	6.67	4.43
Wood	80.	16.67	3.33	5.17
Creaking Wood	90.	6.67	3.33	6.14
Metal	96.67	3.33	0.	6.33
Carpet	66.67	23.33	10.	4.78
Puddles	33.33	66.67	0.	5.8
Water	86.67	13.33	0.	6.57

Table 3: Results of experiment 3: recognition (in percentage) of the surfaces with the recorded real sounds.

found that for the sound of the wood and metal floors more than one participant commented that he/she felt like wearing a different kind of shoe, and for the precision with a solid soil.

In general, the use of the interactive system facilitated the recognition task, and the sound quality of the system was perceived as higher.

One peculiar element of the interactive system is the lack of haptic feedback which is present when walking in the real world and is an important element in the perception of a surface. This lack will be compensated in future implementations of the system, where haptic feedback will be integrated. Some subjects also commented on the importance of visual feedback, which would have obviously helped in the recognition task.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we introduced a real-time footsteps synthesizer controlled by the user, which works independently from the footwear. This is a feature that distinguishes our prototype from other systems developed with similar goals.

The system was tested in a between-subjects experiment, where it was compared to a recognition task including recorded and synthesized offline sounds. Results show that subjects are able to recognize most of the synthesized surfaces using the interactive system with high accuracy. Similar accuracy can be noticed in the recognition of real recorded footsteps sounds, which is an indication of the success of the proposed algorithms and their control.

The developed system is ready to be integrated in computer games and interactive installations where a user can navigate. The simulations proposed, however, reproduce the act of walking on a flat surface.

On the other hand, real life scenarios include also uphill and downhill movements whose footsteps sounds differ significantly from those produced while walking on a flat surface. Such situations can be incorporated in our synthesis engine, by modifying different parameters of the corresponding sounds such as amplitude and temporal variations.

In future work, we indeed plan to utilize the system in multimodal environments, and include haptic and visual feedback, to understand the role of the different sensorial modalities to enhance sense of immersion and presence in scenarios where walking plays an important role.

7 ACKNOWLEDGMENTS

REFERENCES

[1] J. Adrien. *The missing link: Modal synthesis*. MIT Press Cambridge, MA, USA, 1991.

[2] F. Avanzini and D. Rocchesso. Modeling collision sounds: Non-linear contact force. In *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01)*, pages 61–66, 2001.

[3] A. Benbasat, S. Morris, and J. Paradiso. A wireless modular sensor architecture and its application in on-shoe gait analysis. In *Sensors, 2003. Proceedings of IEEE*, volume 2, 2003.

[4] P. Chueng and P. Marsden. Designing Auditory Spaces to Support Sense of Place: The Role of Expectation. In *CSCW Workshop: The Role of Place in Shaping Virtual Community*. Citeseer, 2002.

[5] P. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

[6] P. Cook. Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds. *Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio*, pages 73–78, 2002.

[7] P. Cook. *Real sound synthesis for interactive applications*. AK Peters, Ltd., 2002.

[8] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single state elastoplastic friction models. *IEEE Transactions on Automatic Control*, 47(5):787–792, 2002.

[9] A. Farnell. Marching onwards: procedural synthetic footsteps for video games and animation . *Proceedings of the Pure Data Convention*, 2007.

[10] F. Fontana and R. Bresin. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. *Proc. Colloquium on Musical Informatics*, pages 109–114, 2003.

[11] J. Freeman and J. Lessiter. Hear there & everywhere: the effects of multi-channel audio on presence. *Proceedings of ICAD 2001*, pages 231–234, 2001.

[12] W. Gaver. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*, 5(1):1–29, 1993.

[13] B. Giordano and S. Mcadams. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America*, 119:1171, 2006.

[14] B. Giordano, S. Mcadams, Y. Visell, J. Cooperstock, H. Yao, and V. Hayward. Non-visual identification of walking grounds. *Journal of the Acoustical Society of America*, 123(5):3412, 2008.
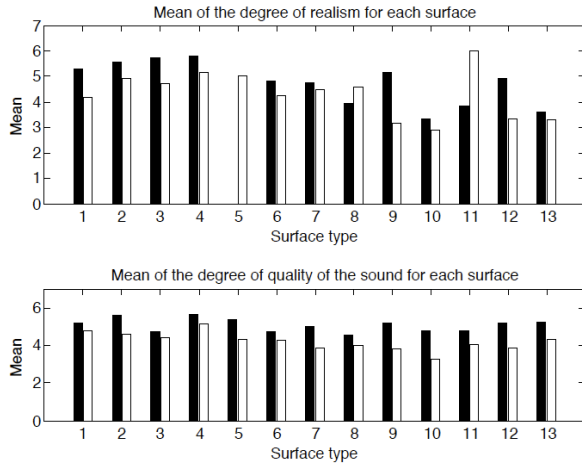
---

[6]www.niwproject.eu

Figure 4: Comparison of the mean of the degree of realism (top) and quality of the sound (bottom) for each surface in experiment 1 (black) and 2 (white). Surface type from left to right: 1-beach sand, 2-gravel, 3-dirt pebbles, 4-snow, 5-high grass, 6-forest underbrush, 7-dry leaves, 8-wood, 9-creaking wood, 10-metal, 11-wood plus reverberation, 12-creaking wood plus reverberation, 13-metal plus reverberation.
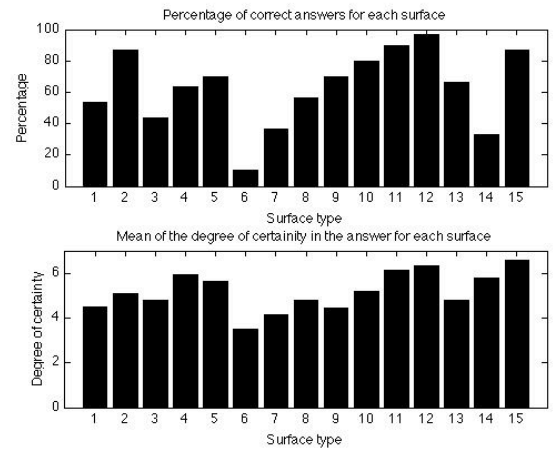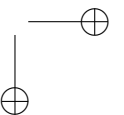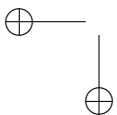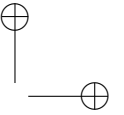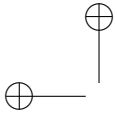
Figure 5: Percentages of correct answers (top) and mean of the degree of certainty in the answer for each surface in experiment 3. Surface type from left to right: 1-beach sand, 2-gravel, 3-dirt pebbles, 4-snow, 5-frozen snow, 6-high grass, 7-forest underbrush, 8-dry leaves, 9-concrete, 10-wood, 11-creaking wood, 12-metal, 13-carpet, 14-puddles and 15-water.

[15] B. Kapralos, D. Zikovitz, M. Jenkin, and L. Harris. Auditory cues in the perception of self-motion. In *116th AES convention*. Citeseer, 2004.

[16] R. Klatzky, D. Pai, and E. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators & Virtual Environments*, 9(4):399–410, 2000.

[17] P. Larsson, D. Västfjäll, and M. Kleiner. Perception of self-motion and presence in auditory virtual environments. In *Proceedings of seventh annual workshop presence*, pages 252–258, 2004.

[18] A. Law, B. Peck, Y. Visell, P. Kry, and J. Cooperstock. A Multimodal Floor-space for Experiencing Material Deformation Underfoot in Virtual Reality. In *IEEE International Workshop on Haptic Audio visual Environments and Games, 2008. HAVE 2008*, pages 126–131, 2008.

[19] X. Li, R. Logan, and R. Pastore. Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America*, 90:3036, 1991.

[20] R. Nordahl. Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion. *Proceedings of Presence*, pages 57–62, 2006.

[21] J. Paradiso, K. Hsiao, and E. Hu. Interactive music for instrumented dancing shoes. In *Proc. of the 1999 International Computer Music Conference*, pages 453–456, 1999.

[22] R. Pastore, J. Flint, J. Gaston, and M. Solomon. Auditory event perception: The source-perception loop for posture in human gait. *Perception and Psychophysics*, 70(1):13, 2008.

[23] N. Raghuvanshi and M. Lin. Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 101–108. ACM New York, NY, USA, 2006.

[24] R. Sanders Jr. *The effect of sound delivery methods on a users sense of presence in a virtual environment*. PhD thesis, NAVAL POSTGRADUATE SCHOOL, 2002.

[25] R. Storms and M. Zyda. Interactions in perceived quality of auditory-visual displays. *Presence: Teleoperators & Virtual Environments*, 9(6):557–580, 2000.

[26] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner. Travelling without moving: Auditory scene cues for translational self-motion. In *Proceedings of ICAD05*, 2005.

[27] K. Van Den Doel, P. Kry, and D. Pai. FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 537–544. ACM New York, NY, USA, 2001.

[28] D. Västfjäll. The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. *CyberPsychology & Behavior*, 6(2):181–188, 2003.

PAPER III

R. Nordahl, S. Serafin and F. Fontana.

Exploring sonic interaction design and presence: Natural
Interactive Walking in Porto

Exploring sonic interaction design and presence:
Natural Interactive Walking in Porto.

Rolf Nordahl, Stefania Serafin
Medialogy, Aalborg University Copenhagen
Lautrupvang 15, 2750 Ballerup, DK
rn,sts@media.aau.dk

Federico Fontana
Dipartimento di Informatica
University of Verona
fontana@sci.univr.it

ABSTRACT

In this paper we report on the results of a three days workshop whose goal was to combine interactive sounds and soundscape design to simulate the sensation of walking in a specific location of Porto.

We discuss advantages and disadvantages of the different solutions proposed in terms of the technology used, and issues of how sonic interaction combined with soundscape design affects presence in virtual environments.

Author Keywords

Sonic interaction, soundscape design, sounds of Porto.

INTRODUCTION

One of the most interesting ways of exploring a city is by walking. From the sonic point of view, walking also allows to express the natural landmark of a place.

When exploring a place by walking, two main categories of sounds can be identified: the person's own footsteps and the surrounding soundscape. Footsteps sounds represent important elements in movies and computer games. In these media, sounds are usually acquired from libraries or recorded by so-called Foley artists that put shoes in their hands and interact with different materials to simulate the act of walking. Recently, several algorithms have been proposed to simulate the sounds of walking algorithmically. One of the pioneers in this field is Perry Cook, who proposed a collection of physically informed stochastic models (PhiSM) simulating several everyday sonic events [3]. Among such algorithms the sounds of people walking on different surfaces were simulated [4]. A similar algorithm was also proposed in [6], where physically informed models simulate several stochastic surfaces.

Recently, in [5] a solution based on granular synthesis was proposed. The characteristic events of a footstep sounds were reproduced by simulating the so-called ground reaction force,i.e., the reaction force supplied by the ground at every step.

Studies on soundscape originated with the work of R. Murray Schafer [11]. Among other ideas, Schafer proposed soundwalks as empirical methods for identifying a soundscape for a specific location. In a soundwalk people are supposed to move in a specific location, noticing all the environmental sounds heard. Schafer claimed that each place has a soundmark,i.e., sounds which one identifies a place with. The idea of experiencing a place by listening has been recently further developed by Blesser and Salter [1]. By synthesizing technical, aesthetical and humanistic considerations, the authors describe the field of aural architecture and its importance in everyday life.

In the field of virtual reality, studies have recently shown how the addition of auditory cues could lead to measurable enhancement in the feeling of presence. Results are available on sound delivery methods [12, 10] or sound quality [2, 10]. Recently, the role of self-sound to enhance sense of presence in virtual environments has been investigated. By combining different kinds of auditory feedback consisting of interactive footsteps sounds created by ego-motion with static soundscapes, it was shown how motion in virtual reality is significantly enhanced when moving sound sources and ego-motion are rendered [7, 8].

The results presented in this paper are part of the Natural Interactive Walking (NIW) FET-Open project[1], whose goal is to provide closed-loop interaction paradigms enabling the transfer of skills that have been previously learned in everyday tasks associated to walking. In the NIW project, several walking scenarios are simulated in a multimodal context, where especially audition and haptics play an important role.

As part of the training sessions of the 2009 Sound and Music Computing Summer school, a workshop was organized. The workshop followed the strategy of combining interactive sounding objects [9] with soundscape design and the role of sound to create a sense of place was addressed. The main goal of the workshop was to allow students to experience the technology developed and to integrate it in the context of the city of Porto.

---

[1]http://www.niwproject.eu/

THE WORKSHOP

The workshop took place in Porto, Portugal, between July 18th and 21st. It lasted four afternoons, and 20 students participated. The first afternoon of the school was dedicated to forming groups. At the end of the afternoon, three students chose the natural interactive walking in Porto workshop. Students were assigned the task to record different soundscapes in Porto, and to combine them with interactive simulations of footsteps sounds as described in the following section. The ultimate goal of the workshop was the recreation in a laboratory setting of the sensation of walking in a specific location of Porto. During the second afternoon the soundscapes of Porto were recorded, as described later. In the third afternoon such soundscapes were combined with the interactive footsteps devices described in the following section. The last afternoon was dedicated to final tuning of the system and demonstrations to the other students.

SONIC INTERACTION DESIGN

When designing an augmented walking surface the physical configuration of the device, the nature of the control system, the sensors and actuators involved, as well as the rendering algorithm need to be taken into account. For example, two basic physical configurations of an augmented walking device can be envisaged (Fig. 1). In our situation, the first configuration consisted of shoes instrumented with sensors attached to them. The second configuration consisted of a rigid surface instrumented with microphones.
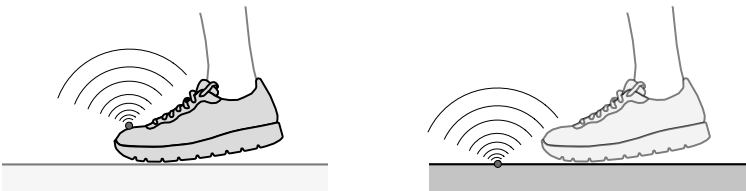

Figure 1. A walking surface augmented using an instrumented shoe (left) or with an instrumented floor (right).

THE FIRST SETUP: SHOES AUGMENTED WITH SENSORS

The first proposed prototype is shown in Figure 2. It provides a technological platform for the development of a multimodal wearable shoe-based interface. The hardware system consists of three main components, respectively providing: sensing, force data acquisition and conditioning, and sound computation and display.

The audio display is provided by two 4Ω Visaton FRWS5 loudspeakers, which were chosen for their low weight and small dimensions. Each speaker is attached onto a shoe, on the surface of the instep. By employing the laptop battery to power the acquisition and computation systems, plus a 9V battery to supply the small loudspeakers, the problem of wearability is addressed. On the other hand, in order to make the system wearable, the loudspeakers have to be small and light, this way raising problems of poor sound quality especially at low frequencies. This problem is addressed by enhancing the system with a portable subwoofer.

Figure 2. A pair of sabots with sensors in the soles. The loudspeakers are fixed to the shoes instep.
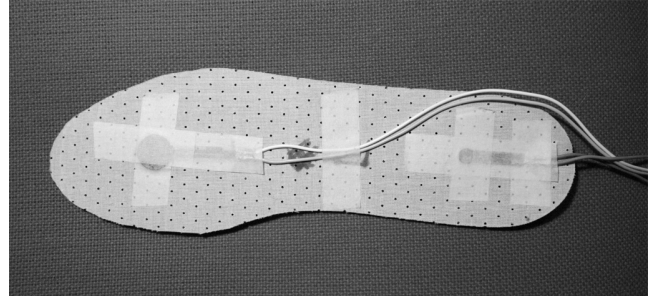

Figure 3. An insole with sensors. In this case, the loudspeaker can be fixed to the ankle with some scotch tape.

Force sensing

Two force sensing resistors (FSRs) have been used in each shoe sole: one for the toe (Interlink[2] 402 FSR) and one for the heel (Interlink 400 FSR). Although these two devices result too sensitive in detecting the weight of a human body, this issue is not a major concern as the sound processing system is enabled by force variations across time rather than absolute force values. The Arduino Duemilanove[3] USB board has been chosen for acquisition purposes.

Usability issues suggest to follow different approaches to make users in condition to wear sensing shoes. For this reason, we have inserted FSRs inside the soles (see Figure 2), as well as put them in the lower side of removable insoles (see Figure 3). The latter solution, in particular, allows to reuse the same sensors in different shoes.

Sound processing and auditory display

The synthesis engine runs on a laptop which computes sounds in real-time according to the algorithm described in Section . The signal so synthesized is sent to a TDA2822 audio power amplifier chip which is supplied by a 9V / 15mAh battery. Both the laptop and the amplifier are placed in a backpack.

Software for sound synthesis

The processing software conditions the sensed force input, and synthesizes sounds. To make it run a on a general-purpose architecture, the software has been written as a com-

[2] www.interlinkelectronics.com
[3] www.arduino.cc

bination of C externals and programs (patches) for the multi-platform real-time DSP environment Pure Data[4] (Pd).

To date, a physically-based crumpling sound model has been tuned to simulate aggregate grounds such as thick snow, gravel, creaking wood.

The model consists of a control layer and a sound synthesis layer implemented as a Pd patch embedding the C externals. More specifically, crumpling sounds are obtained as a result of a stochastic temporal process (control layer) which drives an impact model (sound synthesis), namely triggering impacts with different velocity and at different instants in time. The control layer takes two parameters as input: *force*, which is directly linked to the force data and sets the instantaneous power that enters the overall system, and *resistance*, which determines how resisting the simulated floor materials are, on average, against the incoming force before their state changes. Concerning the physical attributes of the model, parameters of *modal frequencies*, *decay times*, *impact stiffness*, and *shape of the contacting surface* are exposed to control.

Four instances of the model – one for each FSR – are included in separate Pd sub-patches which receive the filtered force data.

THE SECOND SETUP: FLOOR MICROPHONES
We adopted a set of non-contact microphones placed on the floor. In our experiments we used the Shure BETA 91[5], a high performance condenser microphone with a tailored frequency response designed specifically for kick drums and other bass instruments. Its features made it a good candidate for our purpose of capturing the footsteps sounds. In our experiments we placed two microphones on the floor at 1.5 meters distance from each other.

Extraction of the ground reaction force
A footstep sound is the result of multiple micro-impact sounds between the shoe and the floor. The set of such micro-events can be thought as an high level model of impact between an *exciter* (the shoe) and a *resonator* (the floor). In such a vision the sound captured by the microphones can be considered as a composition of both these two components.

In mechanics such exciter is usually called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. The aim of the phase of analysis has been that of finding some parameters that allowed us to extrapolate the exciter from the captured sound, i.e., finding the ground reaction force from the acoustic waveform. Such an extrapolation consisted in removing, from the spectral representation of the sound, the main modes.

In order to achieve the final goal of producing the impression of walking on floors made of different kinds of materials, we removed the contribution of the resonator, kept the exciter and considered the latter as input for a new resonator

that implements different kinds of floors. Subsequently the contribution of the shoe and of the new floor are summed in order to have a complete footstep sound.

The algorithm has been implemented in real-time as an extension to the Pd platform and works as follows: the sound of a person walking is detected in real-time by the microphones described above. From this sound, the resonances corresponding to the impact of the shoe on the floor are removed, in order to extract the GRF. Such GRF is used as input to the sound synthesis algorithms described in the following section.

Sound synthesis and manipulations
The GRF estimated with the technique described in the previous section was used to control two different sound synthesis algorithms, reproducing solid and aggregate surfaces respectively.

To synthesize solid surfaces, we used an algorithm which physically simulates in real-time the contact between hard surfaces. In particular, we controlled one of the input parameters of the model, the impact force, by using the estimated GRF as described in the previous section. This allowed us to recreate realistic footsteps sounds. By varying the different parameters of the model it was possible to simulate the sounds of different surfaces, although a systematic mapping of physical to perceptual parameters is not in place yet.

To synthesize aggregate surfaces, we implemented the physically informed sonic models (PhiSM) [3]. The stochastic energy of the models is controlled by using the estimated GRF.

SOUNDSCAPE RECORDINGS
During the second day of the workshop, the group of students and tutors visited different locations in Porto to record characteristic soundscapes. The tour started from a shopping mall in the outskirt of the city, to the little streets in the central area, one of the main bridges, the metro and the skateboarders playing in front of the Casa da Musica building. All recordings were performed using Zoom H4 or H2 recorders [6]. The students were instructed to record at least three characteristic locations in Porto, and to bring along the recordings the following day.

The third afternoon started with listening the different recordings and discussing why they were interesting and representative of the sounds of Porto. We then encourage the students to produce an overall soundscape to be combined with the interactive footsteps sounds previously described.

PUTTING IT ALL TOGETHER
The integration of the interactive footsteps sounds with the final soundscape was straightforward from the technical point of view, thanks to the capabilities offered by the Pure Data software. In each section of the soundscape, the interactive footsteps sounds were chosen in such a way to match the

---

[4]`www.puredata.info`
[5]http://www.shure.com/

[6]www.zoom.co.jp

particular location. As an example, when the soundscape was recorded at the beach, footsteps sounds of walking on sands were triggered. Due to the limited amount of time, the spatial characteristic of the place were not fine tuned to the corresponding footsteps. In other words, time limitations did not allow to add suitable reverberation effects to the interactive footsteps sounds.

ADVANTAGES AND DISADVANTAGES OF THE TWO SETUPS

Among other things, the workshop provided the possibility to compare the two developed systems and understand their advantages and disadvantages.

Portabilty

Both systems are easily portable. The first system consists of a pair of shoes enhanced with sensors, a subwoofer and a laptop. The second system consists of two microphones, a soundcard and a laptop.

Easiness of setup

To setup the first system it is necessary to ensure that the Arduino is receiving data from the sensors and sending them to the Pure Data environment. Once this step is performed, after a bit of calibration of the sensors data the system is ready to run. The second setup requires the microphone to send data to the Fireface audio interface which sends it to the Pure data environment. So in both situations the easiness of setup stands on the reliability of the interfaces.

Sound quality

The sound quality of the systems obviously depends on the quality of the sound synthesis algorithms as well as the audio delivery methods used. Since the whole setup is intended to reproduce ecological walking sounds resulting from the mechanical interaction between shoes and ground, high sound pressure is not required. Conversely, exceedingly poor low frequencies make the system unsuitable for the rendering of audible low resonances as those elicited when a foot bumps over a hollow floor like a cavity covered with wooden bars. The second setup is delivered through headphones. This is due to the fact that the surrounding sonic environment needs to be relatively quiet, since the microphones should pick up only the real footsteps sounds.

As concerns the quality of the synthesized sounds using the microphone based system, good results have been obtained from the synthesis of big and little gravel, as well as different kinds of wood. In both cases, further improvements to the sound synthesis algorithms are under development.

Sensing capabilities

The shoe enhanced interface forms a solid base where to build further shoe-based functionalities. The reliability of the sensing and conditioning stage, along with the low-latency signal processing and sufficiently realistic auditory display, have already resulted into a credible closed-loop interaction. The speakers and the battery-powered amplifier together give rise to a noticeable lack of low frequencies, neither they are

able to handle sharp transients nor large dynamic ranges. This problem has been addressed by the use of the additional subwoofer. The use of the microphones on the ground did not show any particular problem, and the GRF was extracted in a clear way. The limitation of this system is that only the audio waveform can be obtained, and from that some information extracted in real-time, such as the GRF. As an example, the velocity information can be hardly detected from the acoustic waveform.

However, the envelope extraction as GRF turned out to be the right choice for the control of some parameters of the sound synthesis algorithms.

Wearability

The shoe enhanced system requires users to wear a specific size of footwear, and require a laptop to be carried on the back. On the other end, the floor microphones allow the users to keep their own footwear.

Navigation

The shoe enhanced system ideally allow to navigate without restriction, since the all system is integrated inside the laptop which the users wear. On the other end, the floor microphones require the user to navigate in a specific location delimited by the space inside the microphones.

Integration in VR environments

Both systems have been developed as extension to the Pure Data platform. The platform is open source and can be easily combined with several interfaces and different software packages. A protocol which has been shown to be suitable for integration purposes is the Open Sound Control protocol [7]. In the future, both systems will be integrated with haptic and visual feedback, to simulate different multimodal environments.

CONCLUSIONS AND FUTURE WORK

In conclusion, both systems showed to be suitable as a floor based interaction device to navigate virtual environments. The shoes enhanced with sensors have the advantage of a higher number of sensing capabilities and do not require the environment to be acoustically isolated. On the other end, they require users to be able to wear a particular size of shoe and to wear a backpack with a laptop. The floor microphones have the advantage that users can wear their own shoes. However, they require a quiet environment to be used, and they have limited sensing capabilities.

The integration of the systems with soundscapes of Porto proved to be successful. Although not formally tested, the participants to the workshop enjoyed to be able to virtually walk around the city listening to some characteristic sonic landmarks. Moreover, the students enrolled in the workshop found it valuable to provide a context to the interactive footsteps to create a more imersive experience. The first prototypes of the complete setups could be easily prepared even
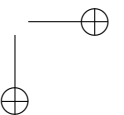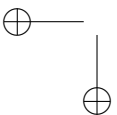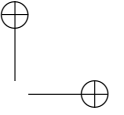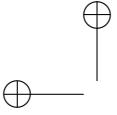
---

[7] opensoundcontrol.org

given the short duration of the workshop. However, their integration can be significantly improved, especially by adding a sense of space to the rendering of the interactive footsteps sounds, and by adding interactivity also to the soundscapes of the city, which right now was simply a static recording.

REFERENCES
1. B. Blesser and L. Salter. *Spaces speak, are you listening?: experiencing aural architecture*. MIT Press, 2006.

2. P. Chueng and P. Marsden. Designing Auditory Spaces to Support Sense of Place: The Role of Expectation. In *CSCW Workshop: The Role of Place in Shaping Virtual Community*. Citeseer, 2002.

3. P. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

4. P. Cook. Modeling Bills Gait: Analysis and Parametric Synthesis of Walking Sounds. *Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio*, pages 73–78, 2002.

5. A. Farnell. Marching onwards: procedural synthetic footsteps for video games and animation. . *Proceedings of the Pure Data Convention*, 2007.

6. F. Fontana and R. Bresin. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In *Proc. Colloquium on Musical Informatics*, pages 109–114, Florence, Italy, May 2003.

7. R. Nordahl. Self-induced Footsteps Sounds in Virtual Reality: Latency, Recognition, Quality and Presence. In *Presence*, 2005.

8. R. Nordahl. Increasing the Motion of Users in Photo-realistic Virtual Environments by Utilising Auditory Rendering of the Environment and Ego-motion. In *Presence*, 2006.

9. D. Rocchesso, R. Bresin, and M. Fernstróm. Sounding objects. *IEEE MULTIMEDIA*, pages 42–52, 2003.

10. R. Sanders Jr. *The effect of sound delivery methods on a users sense of presence in a virtual environment*. PhD thesis, NAVAL POSTGRADUATE SCHOOL, 2002.

11. R. Schafer. *The tuning of the world*. Random House Inc, 1977.

12. R. Storms and M. Zyda. Interactions in perceived quality of auditory-visual displays. *Presence: Teleoperators & Virtual Environments*, 9(6):557–580, 2000.

PAPER IV

Y. Visell, F. Fontana., B. Giordano, **R. Nordahl**, S. Serafin and R. Bresin.

Sound design and perception in walking interactions

ELSEVIER

Sound design and perception in walking interactions

Y. Visell[a,b,*], F. Fontana[c], B.L. Giordano[d], R. Nordahl[e], S. Serafin[e], R. Bresin[f]

[a]*McGill University, Centre for Intelligent Machines and CIRMMT, Montreal, Canada*
[b]*Zurich University of the Arts, Zurich, Switzerland*
[c]*University of Verona, Dipartimento di Informatica, Verona, Italy*
[d]*McGill University, CIRMMT and Schulich School of Music, Montreal, Canada*
[e]*Medialogy, Aalborg University at Copenhagen, Ballerup, Denmark*
[f]*KTH Royal Institute of Technology, CSC School of Computer Science and Communication, Stockholm, Sweden*

Abstract

This paper reviews the state of the art in the display and perception of walking generated sounds and tactile vibrations, and their current and potential future uses in interactive systems. As non-visual information sources that are closely linked to human activities in diverse environments, such signals are capable of communicating about the spaces we traverse and activities we encounter in familiar and intuitive ways. However, in order for them to be effectively employed in human–computer interfaces, significant knowledge is required in areas including the perception of acoustic signatures of walking, and the design, engineering, and evaluation of interfaces that utilize them. Much of this expertise has accumulated in recent years, although many questions remain to be explored. We highlight past work and current research directions in this multidisciplinary area of investigation, and point to potential future trends.
© 2009 Elsevier Ltd. All rights reserved.

Keywords: Auditory display; Vibrotactile display; Interaction design; Walking interfaces

1. Introduction

Just as walking is fundamental to our negotiation of natural environments, it is of increasing relevance to interaction with computational systems. Contact interactions between our feet and the ground play important roles in generating information salient to locomotion control and planning in natural environments, and to the understanding of structures and events in them. Although much of this information is communicated as sound, the latter has been relatively neglected in past research related to walking in human–computer interaction. Consequently, a better understanding of the perception of walking sounds, and the way they may be rendered and displayed, is needed

in order for new and existing human–computer interfaces to effectively make use of these channels. Such developments may hold potential to advance the state of the art in areas such as wearable computers, intelligent environments, and virtual reality. For example, in the ubiquitous computing domain, benefits could be foreseen for a range of new and emerging applications utilizing human locomotion and navigation as means for interaction with digital information (Gaye et al., 2006; Froehlich et al., 2009).

It is important to acknowledge that walking sounds have long played an important role in audiovisual media. In film, footsteps are acknowledged for their ability to signify unseen action, to lend a sense of movement to an otherwise static scene, and to modulate the perception of visible activities. In his seminal work on film sound, Chion writes of footstep sounds as being rich in what he refers to as *materializing sound indices*—those features that can lend concreteness and materiality to what is on-screen, or contrarily, make it seem abstracted and unreal (Chion, 1994). The aim of this paper is to highlight the importance of interdisciplinary research surrounding

---

*Corresponding author at: McGill University, Centre for Intelligent Machines and CIRMMT, Montreal, Canada. Tel.: +1 514 967 1648; fax: +1 418 780 6764.

E-mail addresses: yon@cim.mcgill.ca (Y. Visell), federico.fontana@univr.it (F. Fontana), bruno.giordano@music.mcgill.ca (B.L. Giordano), rn@media.aau.dk (R. Nordahl), sts@imi.aau.dk (S. Serafin), roberto@kth.se (R. Bresin).

sound information in walking for the design of human-interactive systems. In retaining this focus, we address aspects of walking experiences that are seldom investigated in real or virtual contexts. Two potential future scenarios may lend concreteness to the discussion:

- A tourist using a smartphone is able to follow navigational cues that the device supplies by augmenting the sound of his footsteps as if he were walking along a cobblestone trail.
- A search and rescue worker is training in a virtual environment simulation of a dangerous rock canyon area. She receives realistic multimodal cues from the ground surface in the simulator, heightening her sense of immersion.

This article intends to point toward fundamental areas of knowledge needed to effectively realize such applications.

1.1. Foot–ground interactions and their signatures

It is almost a truism to say that self-motion is the most fundamental function of walking. Therefore, it is not surprising that the scientific literature has predominantly attended to questions linked to the biomechanics of human locomotion, and to the systems and processes underlying motor behavior on foot, including the integration of multisensory information subserving planning and control.

Walking is a periodic activity, and a single period is known as the gait cycle. Typical human walking rates are between 75 and 125 steps per minute, corresponding to a fundamental frequency of 1.25–2.08 Hz (Ekimov and Sabatier, 2008). It can be divided into two temporal phases—those of stance and swing. Stance can be characterized in terms of foot position and contact, decomposed into initial heel strike, followed by foot flat, heel off, knee flexion, and toe off (Li et al., 1991). The subsequent swing phase is composed of an initial swing, beginning at toe off. It proceeds to the mid-swing period, when the knee reaches maximum flexion, until the terminal swing, which begins when the tibia is vertical and ends when the reference foot touches the ground. Thus, the gait cycle is characterized by a mixture of postural attributes (e.g., the degree of flexion at the knee) and contact attributes (presence and degree of contact between the plantar area of the foot and the ground). One also distinguishes the several time scales involved, including those of the walking tempo or pace, the individual footstep, encompassing one stance period, and relatively discrete events such as heel strike and toe slap (Fig. 1).

The net force **F** exerted by the foot against the ground can be represented by a time varying spectrum $\mathbf{F}(\omega, t)$,
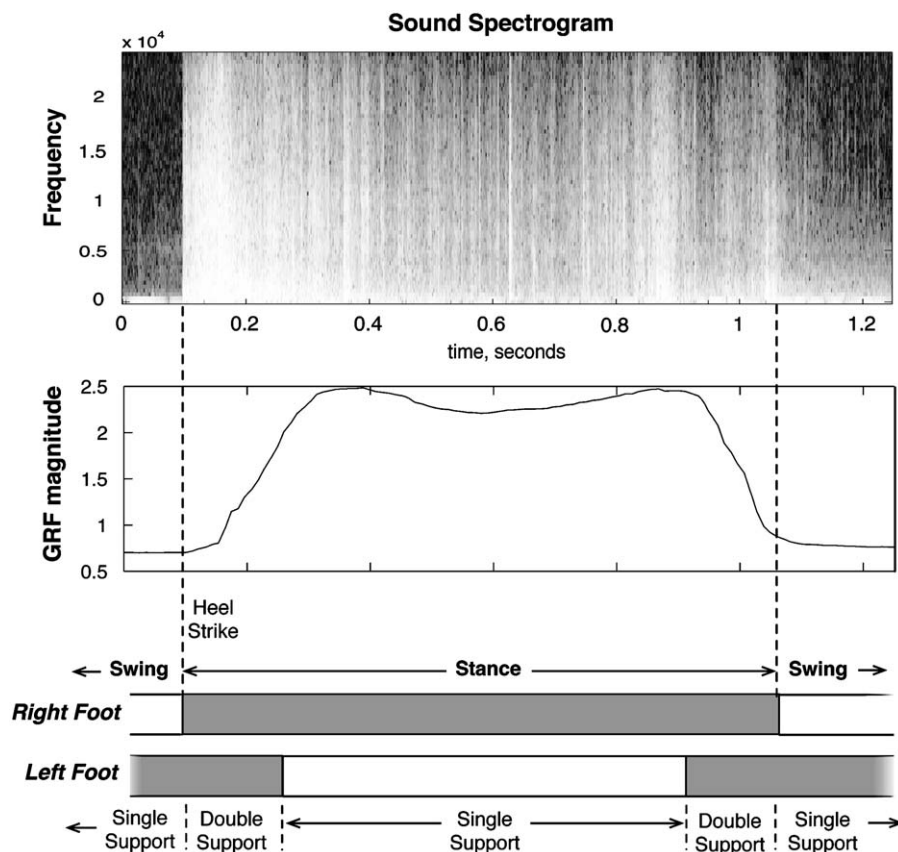

Fig. 1. (Left) Bottom: gait phase for left and right foot, showing more than 50% of the gait cycle. Middle: GRF for right foot (authors' recording). Top: spectrogram showing the acoustic signature resulting from a step onto gravel.

having components tangential and normal to the ground surface; ω denotes angular frequency and t is time. The term ground reaction force (GRF) is often used to refer to the low frequency information in $\mathbf{F}$, below about 300 Hz. The GRF is essentially responsible for the center of mass movement of the individual. It is approximately independent of footwear type, but varies between individuals or walking styles (e.g., Galbraith and Barton, 1970). Higher-frequencies components of $\mathbf{F}(\omega, t)$ can be attributed to fast impacts between heel or toe and ground, sliding friction and contact variations between the shoe and ground (Ekimov and Sabatier, 2006). Unlike the GRF, these components can depend on footwear, on ground surface shape and material properties. They give rise to remote signatures in the form of airborne acoustic signals, seismic vibrations of the ground, and vibrations transmitted through the shoe to the foot, which have been studied in prior literature on acoustic (Li et al., 1991; Ekimov and Sabatier, 2006; Watters, 1965) and vibrational (Cress, 1978; Ekimov and Sabatier, 2008; Galbraith and Barton, 1970) signatures of human walking. These signals vary with the local material and spatial structure of the ground and with the temporal and spatial profile of interactions between the foot of the walker and the ground surface. Several phenomenological models for the contact interactions that produce them are reviewed in Section 3.3.

From a sensory standpoint, in addition to vision, the pedestrian receives sound information via the auditory channel, vibrational information via the tactile (touch) sensory receptors in the skin of the feet, and information about ground shape and compliance via the proprioceptive sense (the body's ability to sense the configuration of its limbs in space). Proprioception, vision, and the vestibular (balance) sense are integrated to inform the pedestrian about his motion in space.

1.2. Overview

As can be seen from the forgoing description, walking generates a great deal of multisensory information about the environment. Prior research has emphasized the influence of visual, haptic, vestibular, and proprioceptive information on control and planning of locomotion over predominantly flat surfaces (e.g., Wu and Chiang, 1996). In two respects, these studies provide a limited account of the complexity of walking in real world environments. Firstly, they have not addressed the range of ground surfaces and materials met outside the lab (e.g., to our knowledge, none has investigated locomotion on soil or gravel). Secondly, they ignore the information contained in sounds generated by walking on real world surfaces (e.g., acoustic information about the gender of a walker, Li et al., 1991). These limitations are addressed in human perception studies presented in Section 2. Notably, in VR contexts, when such layers of perceptual information are available, they are likely to contribute to a heightened sense of presence in the virtual environment, a subject addressed in Section 5.

The remainder of this paper describes developing research on the perception and design of non-visual signatures of walking. We focus on the simulation and perception of foot–ground contact interactions, conceived as carriers of information about ground surfaces and the walkers themselves. In the four sections that follow, we highlight research in these areas:

- The human perception of contact events, with an emphasis on walking sounds.
- Technologies for the interactive synthesis and display of virtual auditory and vibrotactile signatures of walking on natural materials.
- Efficient, physically based computational models for rendering such signals.
- The usability of such displays in human computer interaction and their impact on users' sense of presence in virtual environments.

Their diversity is suggestive of the interdisciplinary effort that is needed to inform future practice in the design of systems that make rich use of walking interactions.

2. Human perception

The information that reaches our senses is structured by the objects and events from which it originates. Probabilistic relationships between the properties of the objects and events in the environment on the one hand, and the structure of the sensory input on the other, are exploited by a perceiver to recover the properties of the surrounding environment. This function of perception is not limited to the visual system, but characterizes all of our senses. In the hearing domain, knowledge has recently accumulated on the perceptual ability to recover properties of the sound generating events in purely acoustical contexts (see Rosenblum, 2004; Lutfi, 2007, for recent reviews).

Locomotion usually produces audible sounds, comprising a number of qualitatively different acoustical events: isolated impulsive signals (e.g., from the impact of a hard heel onto marble); sliding sounds (e.g., a rubber sole sliding on parquet); crushing sounds (e.g., walking on snow); complex temporal patterns of overlapping impulsive signals (e.g., walking on gravel). Overall, the structure of such sounds is jointly determined by several properties of the source: the shape and material of the ground (e.g., brittle ice, gravel), the dynamical features of locomotion itself (e.g., speed, stability), the anthropometric and non-anthropometric properties of a walker (e.g., weight, legs length, but also and gender and emotion of a walker), and the properties of the foot surface in contact with the ground (e.g., size and hardness of the shoe sole). Walking thus conveys information about the properties of the sound source and, even in the absence of explicit training, listeners learn to recover properties of the walking event based on the features of the sound.

There are few published studies on the perceptual processing of walking sounds. Indeed, the major focus in the study of sound source perception has been on impact sounds, impulsive signals generated by a temporally limited interaction between two objects (e.g., mallet hitting a marimba bar). Nonetheless, this literature is relevant to understanding the hearing of walking sounds, for at least two reasons. Firstly, a walking sound is, more often than not, a sequence of isolated impact sounds, and similar strategies are likely applied to recover the properties of the sound source in both cases (e.g., interacting materials). Secondly, theoretical developments in the study of isolated impact sounds (e.g., hypotheses on the nature of inter-individual differences in source perception or on the factors determining the weighting of acoustical information) can, at least in principle, be extended to the perception of any natural sound-generating event.

In Section 2.1 we detail developments on the study of impact sounds. In Section 2.2 we present the literature on the perceptual processing of walking sound events.

2.1. Isolated impact sounds

The study of the perception of isolated impacts is the most developed area within the field of sound source perception (see Giordano et al., in press for a review of prior studies on impact sounds). Typically, research design in this field involves three stages (Li et al., 1991; Pastore et al., 2008). Firstly, the acoustical specification of the properties of the sound source is quantified (e.g., sound frequency is strongly dependent on the size of an object). At times, this analysis aims to quantify the perceptual performance of an ideal listener that perceives a source property through one or more sound features (Pastore et al., 2008; Giordano et al., in press). Secondly, perceptual data are modeled based on mechanical descriptors of the sound source (e.g., McAdams et al., 2004). This stage might consist in a quantification of the accuracy in the human perception of a target source property, or in the analysis of the statistical association between raw behavioral data and all of the manipulated source properties, independent of whether they are the target of perceptual judgment (e.g., material identification is strongly influenced by the size of an object Giordano and McAdams, 2006). Finally, behavioral data are modeled as a function of the sound features. This last modeling stage is of interest to the study of human processing of complex sounds, but also delivers to a sound designer important indications as to those properties of a sound necessary to deliver a perceptual effect.

In the literature on impact sounds, perception of the material of struck objects is linked with energy decay-related properties of the sound signals (e.g., velocity of the loudness decay Giordano and McAdams, 2006); perception of geometrical properties of struck objects is linked with the frequency of the spectral components (e.g., ratios of the frequency of specific spectral components, Lakatos et al.,

1997); perception of the materials of impacting objects is linked with the spectral properties of the early portions of the sounds (e.g., impulsive signals with a high spectral center of gravity are perceived as generated with a hard hammer Giordano et al., in press).

Three recent developments in sound source perception aim at more than quantifying recognition performance and the mechanical and acoustical correlates of perception. Lutfi et al. (2005) investigated the extent to which real sounds can be accurately represented with simplified modal synthesis signals. Experiments compared real and synthetic signals in discrimination and source identification tasks, and investigated discrimination of signals synthesized with a variable number of resonant modes. Results indicate that simplified synthetic sounds, based on a small number of free parameters, yield similar perceptions as their real counterparts, and are frequently indistinguishable from them. Lutfi and Liu (2007) investigated the interindividual variability of the perceptual weight of acoustical information (e.g., the extent to which the frequency of the lowest spectral components affects the perceptual responses). They find that the across-tasks variation of perceptual weights (e.g., the extent to which the perceptual weight of the lowest frequency differs between across the identification of mallet hardness vs. material) is smaller than the across-listeners variation of perceptual weights. They take this result as evidence that participants adopt personalized styles in the weighting of acoustical information, independent of the particular task. They further show that similar performance levels can arise from widely different strategies in the weighting of acoustical information, and that interindividual differences in performance are strongly affected by internal noise factors rather than changes in weighting strategies. Finally, focusing on the estimation of the hardness of impacted objects, Giordano et al. (in press) investigated the influence of the accuracy and exploitability of acoustical information on its perceptual weighting. Studies of source perception reveal that listeners integrate information over both accurate and inaccurate acoustical features, and do not focus selectively onto the most accurate specifiers of a sound source property. It is thus hypothesized that the perceptual weight of an acoustical feature increases with its accuracy and decreases with its perceptual exploitability, as defined by feature-specific discrimination, memory and learning abilities. Both factors appear to interact in determining the weighting of acoustical information. In general, information is weighted in proportion to its accuracy, both in the presence and absence of feedback on response correctness. However, in the absence of feedback the most accurate information can become perceptually secondary, thus signaling limited exploitation abilities.

2.2. Acoustic and multimodal walking events

The study of the human perception of locomotion sounds has addressed several properties of the walking

sound source: the gender (Li et al., 1991; Giordano and Bresin, 2006) and posture of a walker (Pastore et al., 2008), the emotions of a walker (Giordano and Bresin, 2006), the hardness and size of the shoe sole (Giordano and Bresin, 2006), and the ground material (Giordano et al., 2008).

Li et al. (1991) investigated the perception of walkers' gender in untrained listeners. High identification performances were observed. Gender identification appeared related to shoe size, although this factor did not fully account for gender perception. From the acoustical standpoint, gender perception appeared related to the spectral properties of the footstep sounds: females (respectively males) were recognized by shallow (respectively sharp) spectra with a dominant high-frequency (low-frequency) component.

Giordano and Bresin (2006) asked untrained listeners to estimate several properties of the walking sound source: the gender and emotion of a walker (anger, happiness, sadness and fear), and the size and hardness of the shoe soles. The majority of participants recognized each of these attributes at higher-than-chance levels. Interestingly, recognition of gender, sole hardness and size, parameters strongly correlated with each other (female walkers wore smaller shoes with harder soles), was more accurate than the recognition of emotions. Consistent with the results of Li et al. (1991), estimation of gender, and of sole size and hardness, was based on spectral information (again, females were recognized in spectra with a predominant high-frequency component). Perception of emotions was instead strongly influenced by energetic and temporal features: the average pace and pace irregularity, and sound intensity.

Pastore et al. (2008) investigated the discrimination of upright and stooped walking postures in trained listeners. They analyzed the relationship between the mechanics, acoustics and perception of sound events, using the approach described in Section 2.1. The study of the source–acoustics relationship focuses on quantifying the posture discrimination performance afforded by a perceptual focus onto either isolated sound features or onto pairs of sound features (see Giordano et al., in press for a similar analysis conducted with impacted sound sources). They develop a hierarchical model of perceptual decision, based on pairs of sound descriptors. An ideal observer is assumed to be faced with the task of identifying which of two sounds is produced with an upright posture. This observer first considers the difference in the value of an acoustical descriptor between the two sound stimuli. If this difference exceeds a fixed threshold, a response is given. If not, the response is not guessed at random, but is based onto the computation of the difference between the two sound stimuli with respect to a second descriptor. Following this approach in the modeling of a simulated, ideal observer, recognition performance was maximized with pace as the first feature, and spectral amplitude of the heel impact in the 100–500 Hz range as the second.

Giordano et al. (2008) analyzed unimodal and multisensory non-visual identification of two classes of ground materials: solids (e.g., wood) and aggregates (gravels of different sizes). In the multisensory condition, participants walked blindfolded onto ground samples. In the vibrotactile condition, they were also presented with an acoustical masker over wireless headphones. In a proprioception condition, they were presented both the acoustical masker and a tactile masker, delivered through vibrotactile actuators mounted at the bottom of the shoe sole. In the auditory condition, participants did not walk on the materials, but were presented with their own footstep sounds. Overall, identification performance was at chance level only for solid materials in the proprioception condition: absent both auditory and vibrotactile information, solid materials could not be identified. The availability of all sources of non-visual information led to a small but consistent improvement in discrimination performance for solid materials. With aggregates, identification performance was best in the vibrotactile condition, and worst in the auditory condition. Discrimination in the multisensory condition was impaired, compared to that observed in the vibrotactile condition. Limited to the aggregate materials investigated, this result was interpreted as indicating the multisensory integration of incoherent information: auditory on the one hand, vibrotactile and proprioceptive on the other.

3. Augmented ground surfaces as walking interfaces

As noted in the preceding section, the identity of a natural ground surface that is walked upon may be communicated through several different non-visual sensory channels, including auditory, tactile, and proprioceptive channels. Furthermore, material identity may be preserved even when some modalities are absent. Consequently, one way to view the problem of designing an interface for non-visual walking signatures is as a tradeoff between the number of modalities addressed and fidelity at which they can be reproduced, versus the cost and effort required to do so.

One category of applications for the display of non-visual walking signatures in HCI aims to enable walking as a means of controlling self-motion in an immersive virtual environment (VE). In such a context, the convincing representation of ground surface properties is desirable, toward improving a user's sense of presence in the virtual environment (see Section 5). Another category can be identified with systems that utilize walking as a means of interaction with non-immersive systems. For example, such an interface may be designed to enable the use of walking sounds to provide navigational cues (as in the example in the Introduction), or to generate multimedia content (e.g., the PholieMat, described below).

Although somewhat orthogonal to the main content of this paper, we note that considerable research has been undertaken on robotic interfaces for walking in virtual

environments. This subject was recently reviewed by Iwata (2008) and Hollerbach (2008). The devices concerned consist of force-feedback interfaces that, when combined with a virtual environment (VE) simulator, provide the illusion that one is walking in a VE, when one is, in fact, staying in place. One type of configuration for such a device involves an omnidirectional treadmill interface, consisting of a belt that moves under the feet in such a way that the walker remains in place as one walks. Another consists of a pair of platforms attached to the feet and connected to a robotic mechanism capable of delivering forces to the feet. Although such devices are able to approximate the kinesthetic experience of walking (i.e., the sensory experience of bodily motion), it is important to note that they involve an intrinsic cue conflict between the inertial (vestibular) sensory capacities of the body and the visual and kinesthetic cues supplied by the device. Moreover, such devices do not attempt to represent the high frequency tactile or acoustic properties of a surface being walked upon.

The latter properties are the focus of the types of display described here. They consist of walking surfaces augmented with synthetic auditory and/or vibrotactile signals simulating these components of the response of real ground materials (Visell et al., 2007, 2008; Nordahl, 2006). Such devices (e.g., Fig. 5) attempt to compensate for the feedback channels they cannot display—specifically, the felt compliance and shape of the ground—by providing approximately realistic tactile and auditory feedback that is closely coordinated with the footsteps of their users. As indicated in examples presented in Section 5, in a virtual environment context, coordinated visual feedback via wall and/or floor surfaces can also be supplied. For the moment, we concentrate on the interactive auditory and tactile display elements.

A reasonable person may question whether the experience of walking on complex ground materials of varying material, such as marble, earth or snow, can possibly be simulated by a flat ground surface. However, the results of Giordano et al. related above (Giordano et al., 2008) indicate that for solid ground surfaces, vibrotactile and auditory signals are likely more important as carriers of information about material identity than proprioceptive information is. While proprioceptive information is very relevant for the identification of highly compliant materials, the same study suggests that the identity of such materials may be preserved to an acceptable level of accuracy without it. However, further research is needed on the effectiveness of such synthetic information channels at communicating ground properties.

3.1. Physical interaction design

The main components to be specified in the design of an augmented walking surface include the physical embodiment of the device, the sensors and actuators to be employed, and associated electronics for signal acquisition, amplification, and conditioning.

Two basic physical configurations of an augmented walking device can be envisaged (Fig. 2). The reader can undoubtedly envision a number of other possibilities combining these scenarios. The first type consists of a rigid surface instrumented with sensors and actuators. Users are allowed to walk on it wearing ordinary shoes. Such a surface might consist of a flat floor or an isolated surface, such as a stair step. The second type involves a shoe instrumented with sensors integrated in the sole or insole. Portable acoustic actuation can be supplied by a wearable 3D (binaural) auditory display or by wearable loudspeakers. Vibrotactile actuation can be accomplished with actuators integrated within a shoe sole or insole. To date, there has been limited research on such footwear (e.g., the stimulus masking shoes of Giordano et al., 2008), but the technologies involved lie within reach of the state of the art (Hayward and Maclean, 2007). Footwear type and material are relevant in both cases, because natural walking sounds depend on properties of both the shoe and ground (e.g., Li et al., 1991). However, such factors may be best considered in a case-based discussion, as the extent to which user footwear may be known or controlled likely depends upon the application scenario (e.g., virtual environment display vs. augmented reality display in a public space).

The most direct method of sensing involves the acquisition of foot–floor forces or contact regions. Other techniques involve the capture of secondary signatures of foot–ground contact, such as accelerations in a shoe sole or floor surface. A wide range of sensing technologies may be suitable. Examples that have been used for capturing foot–floor contacts include: force-sensing resistive materials (paper, rubber, or other substrates), composite structures of the same type (commercial force sensing resistors), strain gauge based load cells, piezoelectric elements, weaves or films, capacitive elements or films, MEMS accelerometers or gyrometers, and optical fiber composites.

As noted above, auditory display is readily accomplished using standard loudspeakers or head mounted auditory displays. Vibrotactile display, if less common, poses broadly similar requirements. It demands actuators with a frequency response overlapping most of the range from 20 to 1000 Hz (the approximate frequency band of greatest acuity of the human vibrotactile sense, Jones and Sarter, 2008). Moreover, a suitable mechanical design of the
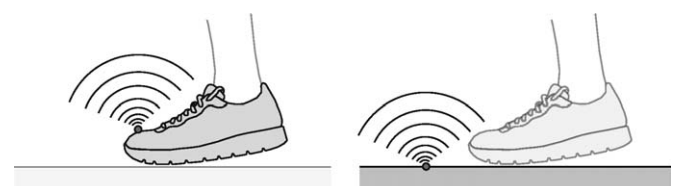

Fig. 2. A walking surface augmented using an instrumented shoe (left) or with an instrumented floor (right).

actuated surface and its structural isolation from the ground is needed to ensure good fidelity and power efficiency. A practical benefit of vibrotactile actuation is that the power requirements are much lower than for a kinesthetic display of comparable scale, in which large forces must be exerted at low frequencies. Among available actuator technologies, linear voice coil actuators, which consist of a magnetic inertial slug suspended on a membrane between a set of electromagnetic coils, are inexpensive, and can be made compact and powerful. Crucially, they permit independent control over stimulus amplitude and waveform. More detailed discussion of tactile actuator types can be found in recent literature (Hayward and Maclean, 2007; Visell et al., 2009).

The spatial distribution of the active display components is another salient factor. If a step is taken on any ground material, contact interactions occur at many sites along the shoe sole. This suggests a high spatial density of sensors and actuators may be required. However, limitations in spatial resolution of the display may be compensated if the interface is designed in such a way that different areas of the foot receive feedback in proportion to the force they are exerting against the tile. This is the case, for example, if the foot receives tactile feedback from a rigid floor surface in response to the force applied to that surface. The proportion may be interpreted as a measure of the responsibility of a given area of the foot for generating the feedback in question.

Commensurate with the coarse spatial resolution of the display, as noted below, for synthesis, a lumped interaction model is frequently adopted, in which the interaction is viewed as taking place through time without spatially distributed degrees of freedom. In such a case, all that may be required is a measurement of the net force applied by each foot at each instant in time. This can be accomplished with a network of sensors with a linear spatial resolution of approximately 30 cm, sufficient to distinguish the net force exerted by each foot.

3.2. Control design

The active components of the display consist of force sensors, actuators and drive electronics, and a computer running the control and sound and/or vibrotactile synthesis algorithm. The control mapping permits user actions captured through the device to determine the synthesis of sounds and/or vibrations.

A simplifying model regards the control mapping as an open loop (Fig. 3), to be calculated independently from the resulting output signals. Such an approximation is tantamount to the segregation of low-frequency input forces (generated by movements of the walker's lower appendages) from higher frequency acoustic and vibrotactile outputs (generated by material interaction with the ground). As described in the Introduction (see Fig. 1), such a separation is supported by prior literature characterizing the information content in comparable signals during walking on real materials (Ekimov and Sabatier, 2006, 2008).

3.3. Sound synthesis

Acoustic and vibrational signatures of locomotion are the result of more elementary physical interactions, including impacts, friction, or fracture events, between objects with certain material properties (hardness, density, etc.) and shapes. The decomposition of complex everyday sound phenomena in terms of more elementary ones has been an organizing idea in auditory display research during recent decades (Gaver, 1993). For present purposes, it is useful to draw a primary distinction between solid and aggregate ground surfaces, the latter being assumed to possess a granular structure, such as that of gravel.

A comprehensive phenomenology of footstep sounds accounting for diverse walking situations should consider various factors, including those described in Section 2.2. Ideally, a designer should have access to a sonic palette making it possible to manipulate all such parameters, including material properties, gestural, and emotional nuances of gait. While this is not yet possible, as reviewed below, there is much prior work on the synthesis of the sounds of contacting objects, including walking settings. Additionally, Section 4 reviews prior work on the control of walking sounds with emotional and gender-based parameters.

3.3.1. Solid surfaces

Sonic interactions between solid surfaces have been extensively investigated, and results are available which describe the relationship between physical and perceptual parameters of objects in contact (Klatzky et al., 2000; van den Doel et al., 2001). Such sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay.

A common approach to synthesizing such sounds is based on a lumped source-filter model, in which an impulsive excitation $s(t)$, modeling the physics of contact, is passed through a linear filter $h(t)$, modeling the response
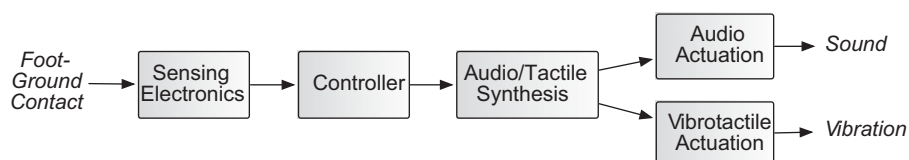

Fig. 3. Footstep interaction viewed as an open-loop process.

of the vibrating object as $y(t) = s(t) \star h(t)$, where $\star$ denotes convolution in time. Modal synthesis (Adrien, 1991) is one widely adopted implementation of this idea. It decomposes of the response model $h(t)$ in terms of the resonant frequencies f_i of the vibrating object (the modes). The response is modeled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, determined by a set of amplitudes a_i, decay rates b_i, and frequencies f_i.

Impacts between shoe and ground (for example, those occurring at heel strike and toe slap) provide the excitation source, while the resonator encompasses either or both of the floor surface itself and the shoe sole. The excitation corresponding to each impact $s(t)$ is assumed to possess a short temporal extent and an unbiased frequency response. In the simplest case, it can be taken to be a known, impulsive signal with total energy E. In a more refined approach, it may consist of a discrete-time model of the force between the two bodies, dependent on additional parameters governing the elasticity of the materials, their velocity of impact, and masses. The parameters governing such solid interactions can be used to specify the characteristics of each impact event, encoding the materials and other interaction parameters, for synthesis using existing models (Avanzini and Rocchesso, 2001).

3.3.2. Aggregate surfaces

The approach outlined above is not directly applicable to cases in which the ground surface does not consist of a solid body. Instead, footsteps onto aggregate ground materials, such as sand, snow, or ice fragments, belie a common temporal process originating with the transition toward a minimum-energy configuration of an ensemble of microscopic systems, by way of a sequence of transient events. The latter are characterized by energies and transition times that depend on the characteristics of the system and the amount of power it absorbs while changing configuration. They dynamically capture macroscopic information about the resulting composite system through time (Fontana and Bresin, 2003).

Physics provides a general formalization of such sounds in terms of: (i) the probabilistic distribution of the energies E of the short transients, which can be assumed to follow a power law $p(E) \propto E^\gamma$. The value of γ determines the type of noise produced by the process (for instance, in the case of crumpling paper it is $-1.6 < \gamma < -1.3$) (Sethna and Dahmen, 2001) and (ii) a model of the temporal density $N(t)$ of transients as a stationary Poisson process, under the assumption that the inter-transient event times τ are assumed to be independent (Papoulis, 1984): $P(\tau) = \lambda e^{\lambda \tau}$.

The parameters γ and λ together determine the macroscopic process dynamics. A simple view of this process is that each transient event consists of a microscopic solid impact with energy E. Thus, in addition, an individual transient can be assumed to possess a resonant response $h(t)$, which is specified in the same way as described above. The resulting parameters characterize each transient event independently of the evolution of the macroscopic system.
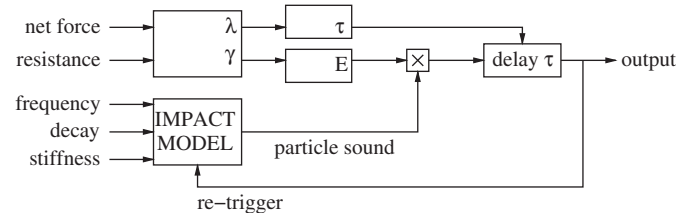

Fig. 4. Algorithm for the generation of sounds of contact with aggregate materials.

Taken together, intensity, arrival times, and impact parameters form a powerful set of independent parametric controls capable of rendering both the process dynamics, which is related to the temporal granularity of the interaction (and linked to the size of the foot, the walking speed, and the walker's weight), and the type of material the aggregate surface is made of (Fig. 4). Such controls enable the sound designer to choose foot–ground contact sounds from a particularly rich physically informed palette. Several models of this general type have been developed in order to mimic the sound of a footstep onto aggregate grounds (Cook, 2002; Fontana and Bresin, 2003; O'Modhrain and Essl, 2004). Section 3.5 reviews one case study in detail.

3.4. Augmented ground surfaces developed to date

The hardware technologies described above are well within the state of the art. As a result, a number of different augmented floor interfaces have been developed, with the largest application domains comprising artistic creation and entertainment. A comparative review of several floor interfaces that were developed for use in musical control was provided by Miranda and Wanderley (2006). Even more attention has been devoted to the development of distributed, sensing floor surfaces, without the explicit intent of generating sound, aided, in part, by the commercially availability of the necessary sensing technologies.[1]

A smaller number of devices have sought to re-create the experience of walking on virtual ground materials. Closest in spirit to the present contribution, Cook (2002) consists of a force-sensing floor mat used as a controller for the real-time synthesis of footstep sounds generated by walking on different ground surfaces. Nordahl (2006) investigated the integration within a VE of self-generated footstep sounds controlled by a set of instrumented sandals (reviewed in Section 5).

3.5. Example: Eco Tile

The Eco Tile, a floor component aimed at the interactive simulation of natural ground materials, is unique in its integration of force sensing in addition to acoustic and vibrotactile actuation (Visell et al., 2007, 2008). The

---

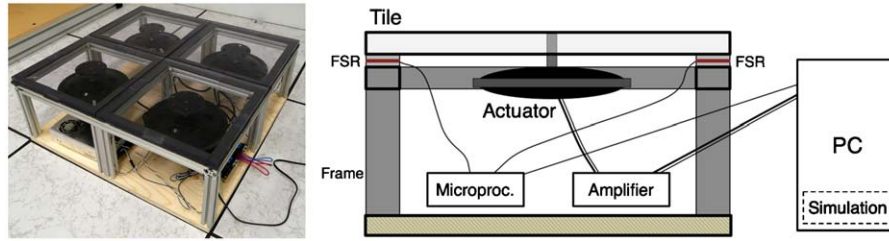[1]Available from, e.g., TekScan, Inc.

Fig. 5. Left: an image of the tile prototype, showing the tile surface (polycarbonate sheet), vibrotactile actuator, force-sensing resistors, structural frame, and associated electronics. Right: diagram of the same, including the PC running the floor material simulation.

prototype shown (Fig. 5) consists of a set of rigid $34 \times 34 \times 0.5$ cm polycarbonate tiles supported by a common structural frame. A linear voice coil actuator (Clark Synthesis model TST239), capable of driving the display over the amplitude and frequency ranges of interest, is rigidly attached to the underside of each tile. Auditory stimuli may be generated in two different ways. If the top of the tile is left untreated, it produces auditory feedback of usable quality as a byproduct of the vibration of the tile surface itself. Alternatively, a separate auditory display may be used. Force sensors are positioned beneath the four corners of each tile, and a single actuator is used to drive the tile surface.

This device has been used to provide the simulation of stepping onto an aggregate ground surface, whose response is synthesized in the manner described in Section 3.3, and controlled by driven by data captured from its sensors as we describe here. Consider a single tile. The vector of four force signals $\mathbf{f}(t)$ from its sensors are used to control the synthesis process. In this case, the distribution of impact events is modeled as a non-homogeneous Poisson random process with a rate parameter $\lambda(t)$ given by

$$\lambda(t) = Au(t)(1 + \tanh(Bu))/2, \qquad (1)$$

$$u(t) = df_L(t)/dt, \quad f_L = \|\mathbf{f}_L\|. \qquad (2)$$

Here A is a control gain parameter, $\mathbf{f}_L(t)$ are components of $\mathbf{f}(t)$ below about 300 Hz, and $(1 + \tanh(Bu))/2$ approximates a Heaviside step function when $B \gg 1$. This simple, force-derivative control scheme guarantees that a response is obtained primarily when the foot is coming down onto the tile, and the force exerted on the tile is increasing (Fig. 6). The total acoustic energy that can be generated by a single footstep is assumed to be a constant[2] value, $\mathscr{E}$. The amount E_i that is attributed to the ith impact event is determined by sampling from an exponential distribution $p(E) \propto E^\gamma$ with free parameter γ, ensuring that $\sum_k E_i = \mathscr{E}$ is satisfied.

Each virtual impact involves an inertial object striking a resonant object with the requisite energy. The force of impact $y(t)$ is determined by a simplified phenomenological equation known as the Hunt and Crossley (1975) model

$$y(t) = kx(t)^\alpha - \lambda x(t)^\alpha \dot{x}(t). \qquad (3)$$

---

[2]For example, $\mathscr{E}$ may be considered to be a constant fraction of the potential energy difference of the body between mid-swing and stance.
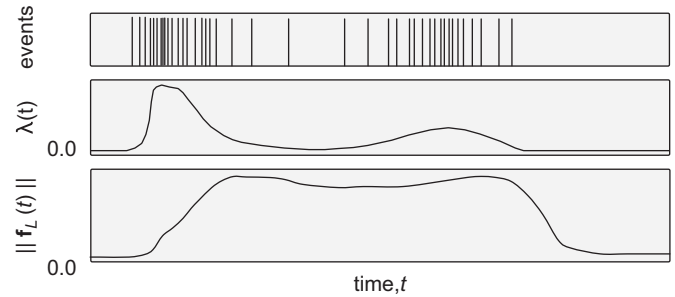

Fig. 6. Qualitative illustration of the control model of Eq. (2), relating the time derivative of the low frequency force signals $d\mathbf{f}_L/dt(t)$ (Bottom) to the event rate parameter $\lambda(t)$ (Middle) and a sampled event sequence (Top).

Here, $x(t)$ is the compression displacement and $\dot{x}(t)$ is the compression velocity. The impact force has parameters governing stiffness k, dissipation λ, and contact shape α. This force is coupled to a modal synthesis representation of the resonant object having the same structure as described above. An impact event is synthesized by initializing Eq. (3) with the velocity v_I of impact and integrating the composite system in time. See Rocchesso and Fontana (2003) for a more detailed discussion. Values for several of the synthesis and control parameters are obtained by measurement and analysis of measured responses of footsteps onto real granular materials (Visell et al., 2008).

In summary, as discussed at the beginning of this section, floor interfaces like the Eco Tile depend for their success on their ability to sustain two distinct illusions: First, that the foot is in contact with a compliant and/or composite material of definite properties that are distinct from those of the floor tile itself; Second, that the virtual physical interaction is distributed across the ground under the foot, rather than originating in a vibration of the ground surface that is (piecewise) constant across the latter.

4. Affective footstep sounds

In this section we present the main results of a recent study in which a model for the synthesis of natural footstep sounds was developed (DeWitt and Bresin, 2007), and preliminarily assessed. The starting point was the model of natural walking and running footstep sounds on aggregate materials that presented in Section 3.3. The pace of footsteps was controlled by tempo curves which were derived from studies in music performance, since strong

similarities between locomotion and music performance were found in prior research. A listening test for the validation of that model highlighted the way in which footstep sequences that were generated using expressive tempo curves, derived from music performance, were perceived as more natural by listeners compared to sequences having a constant pace. Using this study as starting point, we have developed a model of footstep sounds for simulating the presence of people walking in a virtual environment. The design choice was that the footstep sounds should communicate the gender, age, weight, and emotional intention of a virtual walker.

The sound synthesis model was tuned by ear to simulate different ground materials. Gravel, dirt, soft wood, snow, and grass-like settings were selected using the parameters λ and γ; in parallel, the impact parameters were set to reproduce rubber, glass, steel, and wood. The timing in footstep sequences was controlled by using a footstep tempo model developed after measurements of real walkers, who were asked to walk with emotional intentions (happiness, sadness, fear and anger), as well as with their natural (neutral) pace. In interactive listening tests, subjects could adjust pace and material to determine the gender of a virtual walker.

Results show that subjects associated both different pace and material to the two genders (Fig. 7). Female walkers were identified by faster pace (the time interval between to footsteps was about 0.5 s for females and 0.8 s for males), higher resonant frequency for impacts (glass and steel sounds for female; rubber and wood sounds for males) and for particle sounds (mainly gravel and snow sounds for females; dirt and soft wood sounds for males).

It was also tested how subjects would change the emotion of footstep sound sequences. Subjects could control the sound in a 2D activity-valence space in which pace characteristics (regularity and timing) were changed dynamically. Results are promising despite the existence of some confusion between angry and happy footstep sounds. This confusion could be overcome by improving the continuous control over the real-time change of the acoustical characteristics of the ground, thus allowing for a gradually changing perception of both the gender and emotional intention of a virtual walker.

5. VR applications and presence studies

Prior research has addressed issues related to the addition of auditory cues in virtual environments, and whether such cues may lead to a measurable enhancement of immersion in such environments. Most prior work in this area has focused on sound delivery methods (Storms and Zyda, 2000; Sanders and Scorgie, 2002), sound quantity and quality of auditory versus visual information (Chueng and Marsden, 2002) and 3D sound (Freeman and Lessiter, 2001; Vastfjall, 2003). Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in VEs (Larsson et al., 2004; Kapralos et al., 2004; Väljamäe et al., 2005).

Self-generated sounds have been often used as enhancements to VEs and first-person 3D computer games—particularly in the form of footstep sounds accompanying self-motion or the presence of other virtual humans. A smaller number of examples, such as recent work of Law et al. (2008), have even aimed to provide multimodal cues linked to footstep events in such environments (Fig. 8). However, to our knowledge, the effect of such self-generated sounds on users' sense of presence had not been investigated prior to the authors' research in this area. The combination of physics-based rendering of walking sounds with contact-based sensing, as described in the preceding section, also appears to be novel.

5.1. Auditory feedback and motion

The algorithms described in Section 3.3 provided a basis for an evaluation carried out by the authors Nordahl (2006) on the role of interactive self-generated auditory feedback in virtual environments. The visual environment was reproduced using image based rendering techniques capturing part of the botanical garden in Prague. Physically modeled footstep sounds were controlled in real-time via a custom pair of sandals, enhanced with force sensors, which were worn by the user of the environment. The interest of this study was to understand to what extent the quality of auditory feedback would affect users' behavior in such a virtual reality system, and in particular, how and to what extent such interactive auditory feedback
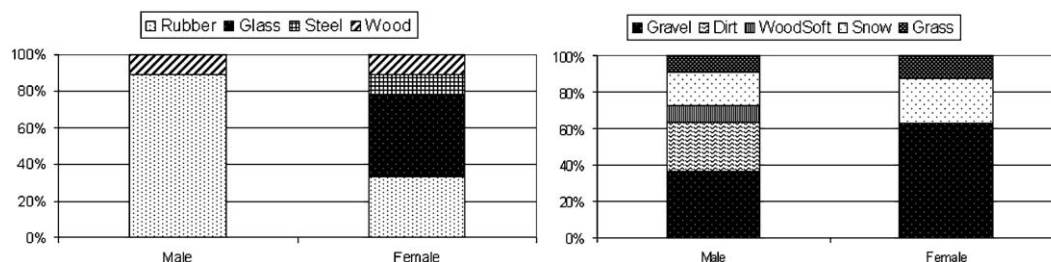

Fig. 7. Subjects' percentage choices of different ground materials in association to walker's gender. The left figure shows the choices for impact sound models. The right figure shows subjects' preferences for different tunings of the crumpling sound model.

Fig. 8. Depiction of the multimodal VE developed by Law et al. (2008), incorporating a CAVE-like visual VE with an auditory and haptic floor display based on the Eco Tile. Footstep events in an immersive snowy landscape are accompanied by visual, auditory, and haptic feedback.

might enhance the motion and presence of subjects in a VE. Prior work on environments simulated using image based rendering techniques has shown that subjects do not find the environments engaging, because of their lack of a dynamic temporal dimension (Turner et al., 2003). The authors were motivated by the belief that interactive auditory feedback can address such limitations.

This hypothesis was tested in an experiment with 126 subjects. Before entering the room, subjects were asked to wear a head mounted display and the instrumented sandals. Subjects were not informed about the purpose of the sensor-equipped footwear. Before beginning 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 afterward they would be asked to fill out a questionnaire with several questions focused on what they remembered 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 uni-modal (visual) input.
(2) Visual with footstep sounds. In this condition, subjects had bi-modal perceptual input including auditory feedback with non-self-generated environmental sounds (audio-visual), comparable to earlier research (Nordahl, 2005).
(3) Visual with full sound. In this condition implies subjects were provided with environmental sounds, spatialized footstep sounds (using the VBAP algorithm) as well as rendering sounds from ego-motion (the subjects triggered sounds via their own footsteps).
(4) Visual with full sequenced sound. This condition was strongly related to condition 3. However, it was run in
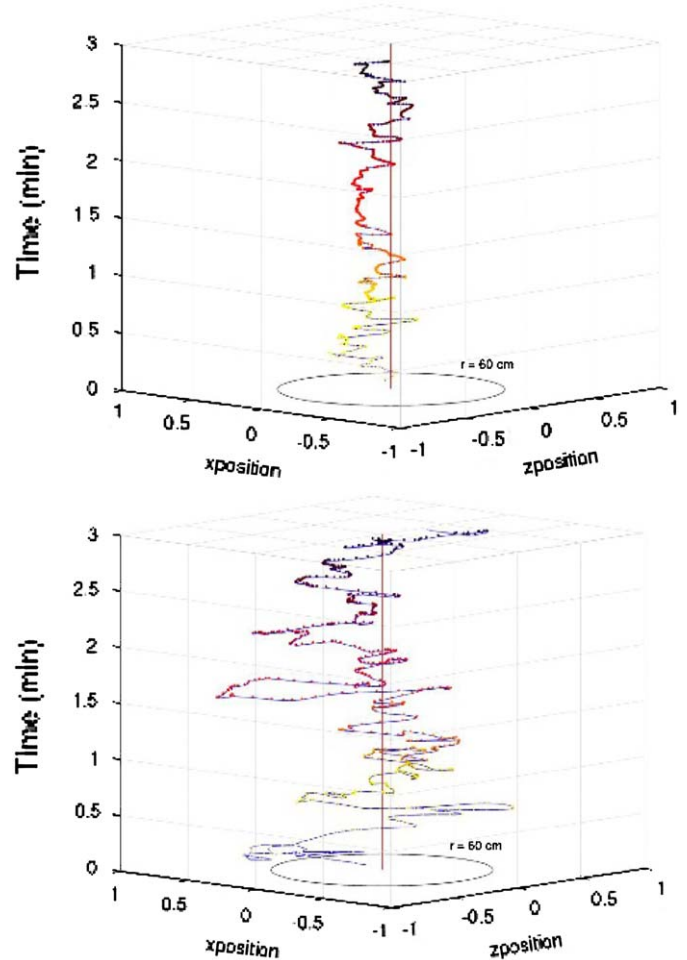

Fig. 9. Top: visualization over time of the motion of one subject over time with visual only condition (top) and full condition (bottom).

three stages: the condition started with bi-modal perceptual input (audio-visual) with static sound design. After 20 s, the rendering of the sounds from emotion was introduced. After 40 s the 3D sound started (in this case the sound of a mosquito, followed by other environmental sounds).
(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 music[3] described before. 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.

The results provided clear indications that footsteps sounds, when combined with environmental sounds,

---

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

significantly enhance the motion of subjects in such a VE. The quantity of motion is clearly visible in Fig. 9, which shows subject position in the 2D plane, as acquired from a Polhemus magnetic tracker placed on the top of the head, respectively for one subject with *Visual only* stimuli (top) and with *Full condition* (bottom). The increase of movement exhibited by this subject in the *Full*condition is clearly noticeable. Results also indicated that footsteps sounds alone do not appear to cause a significant enhancement in subjects' motion. 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 suggests that environmental sounds are implicitly necessary in a VE, and we assume that their inclusion is important to facilitate motion. Further detail is provided in the indicated references.

6. Conclusions

The interactive simulation of acoustic contact signatures generated by walking, which are highly salient to the experience of locomotion in diverse everyday environments, requires solving a number of design, engineering, and evaluation problems that make the realization of such interfaces a complex and multidisciplinary task. To effectively design the feedback channels involved, a solid base of knowledge on the perception of sound and vibrotactile information in walking events is needed, building on those studies discussed in this article. Conversely, we expect such knowledge to be further developed through experiments conducted in both real environments, and through virtual environments utilizing the current state of the art in acoustic and haptic display. The technologies, algorithms and methods for multimodal simulation and evaluation reviewed here are already capable of contributing to this process, but each can be further improved upon. For example, measurements relating the high frequency acoustic response of different ground surfaces to low frequency gait profiles (GRFs) would allow to refine the acoustic rendering techniques described in Section 3.3. Joint measurement of such attributes has only recently been broached in the literature (Visell et al., 2008). Control and rendering models can be refined to match the limitations of display devices and the perceptual capacities of their users, with the aim of compensating, as far as possible, missing sensory channels, such as proprioception. On the side of material attributes, a unification of rendering algorithms might be achieved by more carefully modeling of the physics of interaction with the materials involved, whether in a deterministic or stochastic setting. Such techniques have been successfully used in prior literature on everyday sound synthesis (Rath and Rocchesso, 2005). Open problems such as these can be expected to sustain the vitality of research in this emerging field for many years to come.
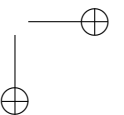
References

Adrien, J., 1991. The missing link: modal synthesis. In: Representations of Musical Signals. MIT Press, Cambridge, MA, USA, pp. 269–298.

Avanzini, F., Rocchesso, D., 2001. Modeling collision sounds: non-linear contact force. In: Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX-01), pp. 61–66.

Chion, M., 1994. Audio-Vision. Columbia University Press, New York, USA.

Chueng, P., Marsden, P., 2002. 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.

Cook, P.R., 2002. Modeling bill's gait: analysis and parametric synthesis of walking sounds. In: Proceedings of the Audio Engineering Society 22 Conference on Virtual, Synthetic and Entertainment Audio, AES, Espoo, Finland.

Cress, D.H., 1978. Terrain considerations and data base development for the design and testing of devices to detect intruder-induced ground motion. Technical Report M-78-1, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

DeWitt, A., Bresin, R., 2007. Sound design for affective interaction. In: Affective Computing and Intelligent Interaction ACII 2007, pp. 523–533.

Ekimov, A., Sabatier, J., 2006. Vibration and sound signatures of human footsteps in buildings. Journal of the Acoustical Society of America 120 (2), 762–768.

Ekimov, A., Sabatier, J., 2008. A review of human signatures in urban environments using acoustic and seismic methods. In: Proceedings of IEEE Technologies for Homeland Security.

Fontana, F., Bresin, R., 2003. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In: Proceedings of the Colloquium on Musical Informatics, Florence, Italy, pp. 109–114.

Freeman, J., Lessiter, J., 2001. Hear, there and everywhere: the effects of multichannel audio on presence. In: Proceedings of ICAD 2001, pp. 231–234.

Froehlich, P., Simon, R., Baille, L., 2009. Mobile spatial interaction special issue. Personal and Ubiquitous Computing, May.

Galbraith, F., Barton, M., 1970. Ground loading from footsteps. Journal of the Acoustical Society of America 48 (5B).

Gaver, W.W., 1993. What in the world do we hear? An ecological approach to auditory source perception. Ecological Psychology 5 (1).

Gaye, L., Holmquist, L.E., Behrendt, F, Tanaka, A., 2006. Mobile music technology: Report on an emerging community. In: Proceedings of NIME'06.

Giordano, B.L., Bresin, R., 2006. Walking and playing: What's the origin of emotional expressiveness in music? In: Baroni, M., Addessi, A.R., Caterina, R., Costa, M. (Eds.), Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9), Bologna, Italy.

Giordano, B.L., McAdams, S., 2006. Material identification of real impact sounds: effects of size variation in steel, glass, wood and plexiglass plates. Journal of the Acoustical Society of America 119 (2), 1171–1181.

Giordano, B.L., McAdams, S., Visell, Y., Cooperstock, J., Yao, H., Hayward, V., 2008. Non-visual identification of walking grounds. Journal of the Acoustical Society of America 123 (5), 3412.

Giordano, B.L., Rocchesso, D., McAdams, S., Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. Journal of Experimental Psychology: Human Perception and Performance, in press.

Hayward, V., Maclean, K., 2007. Do it yourself haptics. Part I. IEEE Robotics and Automation Magazine 14 (4).

Hollerbach, J., 2008. Locomotion interfaces and rendering. In: Lin, M., Otaduy, M. (Eds.), Haptic Rendering: Foundations, Algorithms and Applications. AK Peters, Ltd.

Hunt, K.H., Crossley, F.R.E., 1975. Coefficient of restitution interpreted as damping in vibroimpact. ASME Journal of Applied Mechanics 42 (2), 440–445.

Iwata, H., 2008. Haptic interface. In: Sears, A., Jacko, J.A. (Eds.), The Human–Computer Interaction Handbook, second ed. Lawrence Erlbaum Associates, New York, 2008.

Jones, L.A., Sarter, N.B., 2008. Tactile displays: guidance for their design and application. Human Factors 50 (1), 90–111.

Kapralos, B., Zikovitz, D., Jenkin, M., Harris, L., 2004. Auditory cues in the perception of self-motion. In: Proceedings of the 116th AES convention.

Klatzky, R.L., Pai, D.K., Krotkov, E.P., 2000. Perception of material from contact sounds. Presence: Teleoperators and Virtual Environment 9 (4), 399–410.

Lakatos, S., McAdams, S., Caussé, R., 1997. The representation of auditory source characteristics: simple geometric form. Perception & Psychophysics 59 (8), 1180–1190.

Larsson, P., Västfjäll, D., Kleiner, M., 2004. Perception of self-motion and presence in auditory virtual environments. In: Proceedings of the Seventh Annual Workshop Presence, pp. 252–258.

Law, A.W., Peck, B.V., Visell, Y., Kry, P.G., Cooperstock, J.R., 2008. A multi-modal floor-space for displaying material deformation underfoot in virtual reality. In: Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their Applications.

Li, X., Logan, R.J., Pastore, R.E., 1991. Perception of acoustic source characteristics: walking sounds. Journal of the Acoustical Society of America 90 (6), 3036–3049.

Lutfi, R.A., 2007. Human sound source identification. In: Yost, W.A., Fay, R.R., Popper, A.N. (Eds.), Auditory Perception of Sound Sources. Springer, New York, NY, pp. 13–42.

Lutfi, R.A., Liu, C.J., 2007. Individual differences in source identification from synthesized impact sounds. Journal of the Acoustical Society of America 122, 1017–1028.

Lutfi, R.A., Oh, E., Storm, E., Alexander, J.M., 2005. Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians, and nonmusicians. Journal of the Acoustical Society of America 118, 393.

McAdams, S., Chaigne, A., Roussarie, V., 2004. The psychomechanics of simulated sound sources: material properties of impacted bars. Journal of the Acoustical Society of America 115 (3), 1306–1320.

Miranda, E., Wanderley, M., 2006. New Digital Musical Instruments: Control And Interaction Beyond the Keyboard, AR Editions.

Nordahl, R., 2005. Auditory rendering of self-induced motion in virtual reality. M.Sc. Project Report, Department of Medialogy, Aalborg University Copenhagen.

Nordahl, R., 2006. Increasing the motion of users in photorealistic virtual environments by utilizing auditory rendering of the environment and ego-motion. In: Proceedings of Presence, pp. 57–62.

O'Modhrain, S., Essl, G., 2004. Pebblebox and crumblebag: tactile interfaces for granular synthesis. In: Proceedings of the NIME 2004, Hamamatsu, Japan, pp. 74–79.

Papoulis, A., 1984. Probability, Random Variables, and Stochastic Processes, second ed. McGraw-Hill, New York.

Pastore, R.E., Flint, J.D., Gaston, J.R., Solomon, M.J., 2008. Auditory event perception: the source-perception loop for posture in human gait. Perception & Psychophysics 70 (1), 13–29.

Rath, M., Rocchesso, D., 2005. Continuous sonic feedback from a rolling ball. IEEE Multimedia 12 (2), 60–69.

Rosenblum, L.D., 2004. Perceiving articulatory events: lessons for an ecological psychoacoustics. In: Neuhoff, J.G. (Ed.), Ecological Psychoacoustics. Elsevier Academic Press, San Diego, CA, pp. 219–248.

Rocchesso, D., Fontana, F., 2003. The Sounding Object. Edizioni di Mondo Estremo, Florence, Italy.

Sanders, R., Scorgie, M., 2002. The Effect of Sound Delivery Methods on a User's Sense of Presence in a Virtual Environment.

Sethna, J.P., Dahmen, K.A., 2001. Myers, crackling noise. Nature 410, 242–250.

Storms, R.L., Zyda, M.J., 2000. Interactions in perceived quality of auditory-visual displays. Presence: Teleoperators & Virtual Environments 9 (6), 557–580.

Turner, S., Turner, P., Carroll, F., O'Neill, S., Benyon, D., McCall, R., Smyth, M., 2003. Re-creating the Botanics: towards a sense of place in virtual environments. In: Proceedings of the Environmental Science Conference.

van den Doel, K., Kry, P., Pai, D., 2001. FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 537–544.

Väljamäe, A., Larsson, P., Västfjäll, D., Kleiner, M., 2005. Travelling without moving: Auditory scene cues for translational self-motion. In: Proceedings of ICAD'05.

Vastfjall, D., 2003. The subjective sense of presence, emotion recognition, and experienced emotions in auditory virtual environments. CyberPsychology & Behavior 6 (2), 181–188.

Visell, Y., Cooperstock, J., Franinovic, K., 2007. The ecotile: an architectural platform for audio-haptic simulation in walking. In: Proceedings of the 4th International Conference on Enactive Interfaces.

Visell, Y., Cooperstock, J., Giordano, B.L., Franinovic, K., Law, A., McAdams, S., Jathal, K., Fontana, F., 2008. A vibrotactile device for display of virtual ground materials in walking. In: Proceedings of Eurohaptics 2008.

Visell, Y., Law, A., Cooperstock, J., 2009. Touch is everywhere: floor surfaces as ambient haptic displays. IEEE Transactions on Haptics, 2 (3).

Watters, B.G., 1965. Impact noise characteristics of female hard-heeled foot traffic. Journal of the Acoustical Society of America 37, 619–630.

Wu, G., Chiang, J.H., 1996. The effects of surface compliance on foot pressure in stance. Gait & Posture 4 (2), 122–129.
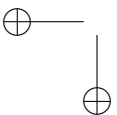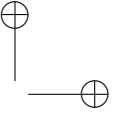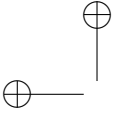
PAPER V

R. Nordahl

Sonic Interaction Design to enhance presence and motion
in virtual environments.

Sonic Interaction Design to enhance presence and motion in virtual environments

Rolf Nordahl
Medialogy, Aalborg University Copenhagen
Lautrupvang 15, 2750
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 since only visual stimulus was provided. By transferring information from film studies and current practice, practitioners emphasize that auditory feedback such as sound of footsteps signifies the character giving them weight and thereby subjecting the audience to interpretation of embodiment.

We hypothesize that the movement rate can be significantly enhanced by introducing auditory feedback. In the described study, 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 subjects' motion in virtual reality is significantly enhanced when dynamic sound sources and sound of egomotion are rendered in the environment.

Author Keywords

Interactive auditory feedback, physical models, presence

ACM Classification Keywords

H 5.5 Sound and Music Computing ; H 5.2 User Interfaces

INTRODUCTION

In the realm of Virtual Reality (VR) and Virtual Environments (VE) sound has not until very recently been considered of value when one wishes to mediate experiences to the participant. 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 [15]. While research has provided different results concerning multimodal interaction among the senses, several questions remain in how one can utilize e.g., audiovisual phenomena when building interactive VR experiences.

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 as immersion [9].

Engineers have been interested in the audio-visual interaction from the perspective of optimizing the perception of quality offered by technologies [6, 14]. Furthermore, studies have shown that by utilizing audio, the perceived quality of lower quality visual displays can increase [16].

Likewise researchers from neuroscience and psychology have been interested in the multimodal perception of the auditory and visual senses [8]. 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 [2] and the ventriloquism effect [7].

Among the initiatives to investigate how technology can enhance sense of immersion in virtual environments, the recently completed BENOGO project[1] had as its main focus the development of new synthetic image rendering technologies (commonly referred to as Image Base Rendering (IBR)) that allowed photo-realistic 3D real-time simulations of real environments. The project aimed at providing a high degree of immersion to subjects for perceptual inspection through artificial created scenarios based on real images. Throughout the project the researchers wished to contribute to a multilevel theory of presence and embodied interaction, defined by three major concepts: immersion, involvement and fidelity.

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 as static over time. Depth perception and direction are varied according to the motion of the user, which is able to investigate the environment at 360^0 inside the so-called region of exploration (REX). However, no events happen in the environment, which make it rather uninteresting to explore [11].

In this paper we advocate the use of interactive auditory feedback as a mean to enhance immersion in a photorealistic virtual environment. We focus both on ambient sounds, defined as sounds characteristic of a specific environment which the user cannot modify, as well as interactive sounds, which were synthesized in real-time and controlled by actions of users in the environment. Such sounds were driven by using a footsteps controller able to capture the motion of

---

[1] www.benogo.dk

users in the environment, and subsequently produce in real-time sounds produced by walking on different surfaces.

Furthermore, sounds were spatialized in a 8-channels surround sound system utilizing the Vector Based Amplitude Panning (VBAP) algorithm [12]. Our hypothesis is that augmenting the environment with interactive sounds will enhance motion of the subjects. Measuring the quantity of motion is important since we hypothesize that a higher level of motion implies that subjects explore the environment more actively and therefore with an increased interest.

A MULTIMODAL ARCHITECTURE

Figure 1 shows a schematic representation of the multimodal architecture used for the experiments. The visual stimulus was provided by a standard PC running Suse Linux 10. This computer was running the 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: ((640x3)x480), (921,600 color elements) equivalent to 307,200 triads. Furthermore the HMD provides a field of view of 60^0 diagonal. The tracker used was a Polhemus IsoTrak II3. It 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[2] and as output module a Fireface 800 from RME5[3] was used. Sound was delivered by eight Dynaudio BM5A speakers[4]. Figure 2 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.

AUDITORY RENDERING

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 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[5], where eight channels surround is presently implemented. This is the reason why 8-channels sound rendering was chosen compared to e.g., binaural rendering [3].

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

---

[2]www.cycling74.com

[3]http://www.rmeaudio.com/english/firewire/
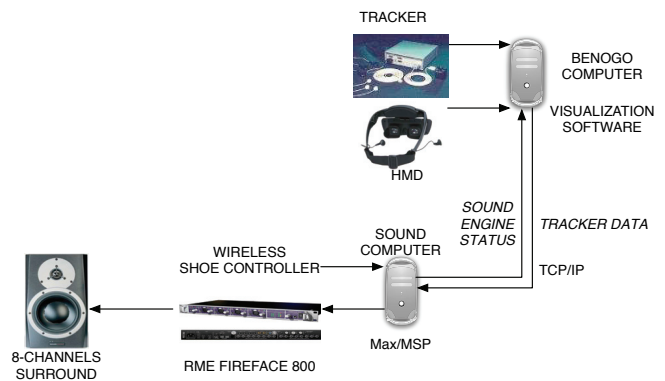
[4]http://www.dynaudioacoustics.com

Figure 1. Connection of the different hardware components in the experimental setup.

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.

Four kinds of auditory feedback 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, 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 [11])

4. A piece of classic music as described before, reproduced at max. peak of 58dB, measured c-weighted with slow response.

However, six testing conditions were implemented, as described later.

Interactive auditory feedback

A real-time footstep synthesizer, controlled by the subjects using a set of sandals embedded with pressure sensitive sensors was designed. By navigating in the environment, the user controlled synthetic sounds. Footsteps recorded on seven different surfaces were obtained from the Hollywood Edge Sound Effects library.[5] The surfaces used were metal, wood, grass, bricks, tiles, gravel and snow. The sounds were analyzed and the analysis results used to build a footsteps synthesizer using a combination of modal synthesis [1] and phys-

---

[5]www.hollywoodedge.com

Figure 2. A view of the 8-channels surround sound lab where the experiments were run.

Figure 3. The sandals enhanced with pressure sensitive sensors wirelessly connected to a microprocessor.

ically informed stochastic models (PHYSM) [4]. More specifically, regular surfaces such as bricks, metal, wood and tiles were synthesized using a modal synthesizer with few (two or three) resonances. Grass, gravel and snow were synthesized using the PHYSM algorithm. In order to control the synthetic footsteps in the virtual reality environment, subjects were asked to wear a pair of sandals embedded with pressure sensitive sensors placed one in each heel. Such sandals are shown in Figure 3.

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 Max/MSP.[6] The sensors were wirelessly connected to a microprocessor, as shown in Figure 3, and the microprocessor was connected to a laptop PC.

The continuos pressure value was used to control the force of the impact of each foot on the floor, to vary the quality of the synthetic 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. The different simulated surfaces were activated according to the virtual place which the users were visiting, and rendered through and 8-channel surround sound system.

VISUAL FEEDBACK
The visual feedback used in these experiments was created under the BENOGO project. The idea behind this project is the creation of photorealistic visual environments obtained by taking pictures of a specific location at different angles,

---

[6]www.cycling74.com

and building a reconstruction of the same place at the computer using image based rendering techniques. In this specific experiment, subjects were looking at pictures from the Prague botanical garden, which is shown in Figure 4.

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.

Figure 4. An image of the Prague botanical garden used as visual feedback in the experiments.

TEST DESCRIPTION

Figure 5. A subject navigating in the virtual environment wearing an head mounted display HMD)

126 subjects took part to the experiment. All subjects reported normal hearing and visual conditions. Figure 5 shows one of the subjects participating to 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 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.

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 [11].

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 the VBAP algorithm) 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 egomotion was introduced. After 40 seconds the 3D sound started (in this case the sound of a mosquito, followed by other environmental sounds).

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 music[7] described before. 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.

CONDITION	AUDITORY STIMULI	NUM SUBJ	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 six different conditions to which subjects were exposed during the experiments. The number in the second column refers to the auditory feedback previously described.

RESULTS

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

Table 2: Motion analysis for the different conditions.

	Visual only	Visual w. foot	Full	Full seq.	Music
Visual only			0.006	0.03	0.41
Visual w. foot	0.26		0.04	0.132	0.197
Full					
Full seq.			0.243		0.018
Sound + 3D	0.431	0.32	0.022	0.048	0.347
Music			0.003		

Table 3: Comparison of the motion analysis for the different conditions (p-value).

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

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. 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, and not additional body motion.

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. 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 *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 motion.

Figure 6 shows the visualization of the Polhemus tracker data for one subject with visual only stimuli (top) and with full condition (bottom). The increase of movement in the full condition is clearly noticeable.

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 [17]. The SVUP is concerned with examining 4 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 4 questions which all relates 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 related 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 [10], from 1-7, (from 1 represents not at all, and 7 represents very much).
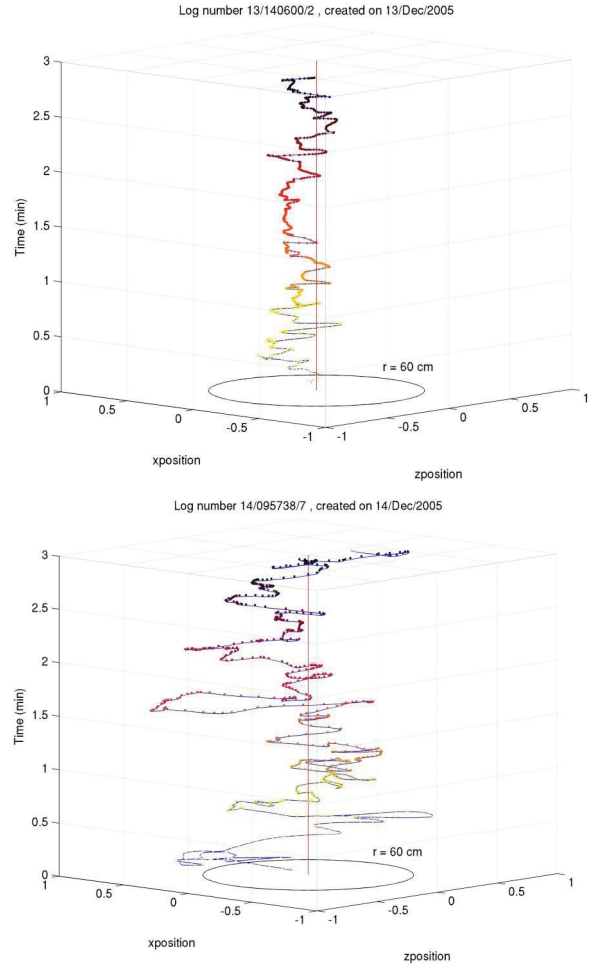

Figure 6. Top: visualization of the motion of one subject over time with visual only condition (top) and full condition (bottom).

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

Table 4: Average presence index for the 6 experimental conditions.

Table 4 shows the results of the presence questionnaire for the different conditions. As can be seen from the Table, 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 thesis 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, i.e., 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 that such experiments.

CONCLUSION

In this paper we investigated the role of dynamic sounds in enhancing motion and presence in virtual reality. Results show that 3D sound with moving sound sources and auditory rendering of ego-motion significantly enhance the quantity of motion of subjects visiting the VR environment.
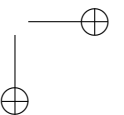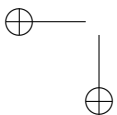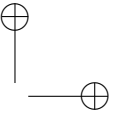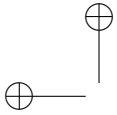
It is very interesting to notice that it is not the individual auditory stimulus that affects the increase of motion of the subjects, but rather that it is the combination of soundscapes, 3-dimensional sound and auditory rendering of ones own motion that induces a higher degree of motion.

We also investigated if 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 were the visual feedback is more dynamic and interactive, such as computer games and virtual environments reproduced using 3D graphics.

REFERENCES

1. J.M. Adrien. The missing link: modal synthesis. *Representations of musical signals table of contents*, pages 269–298, 1991.

2. B. Arons. A review of the cocktail party effect. *Journal of the American Voice*, 12, 1992.

3. D.R. Begault. *3-D sound for virtual reality and multimedia*. AP Professional Boston, 1994.

4. P.R. Cook. Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds. *Computer Music Journal*, 21(3):38–49, 1997.

5. C. Cruz-Neira, D.J. Sandin, T.A. DeFanti, R.V. Kenyon, and J.C. Hart. The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992.

6. N.F. Dixon and L. Spitz. The detection of audiovisual desynchrony. *Perception*, 9, 1980.

7. S. Handel and I. NetLibrary. *Perceptual Coherence Hearing and Seeing*. Oxford University Press, 2006.

8. A Kohlrausch and S. Vand de. Par. Auditory-visual interaction: From fundamental research in cognitive psychology to (possible) applications. In *Proceeding of IST/SPIE Conference on Human Vision and Electronic Imaging IV*, 1999.

9. M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):20, 1997.

10. TJ MAURER and HR PIERCE. A comparison of Likert scale and traditional measures of self-efficacy. *Journal of applied psychology*, 83(2):324–329, 1998.

11. R. Nordahl. Auditory rendering of self-induced motion in virtual reality. *M. Sc. project report, Dept. of Medialogy, Aalborg University Copenhagen*, 2005.

12. V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.

13. V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Helsinki University of Technology, 2001.

14. 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.

15. K.M. Stanney et al. *Handbook of virtual environments: design, implementation, and applications*. Lawrence Erlbaum Associates, 2002.

16. R.L. Storms and M.J. Zyda. Interactions in perceived quality of auditory visual displays. *Presence*, 9, 2000.

17. D. Västfjäll, P. Larsson, and M. Kleiner. Development and validation of the Swedish viewer-user presence questionnaire (SVUP), 2000.

PAPER VI

R. Nordahl, and D. Korsgaard,

Distraction as a measure of presence: using visual and
tactile adjustable distraction as a measure to determine
immersive presence of content in mediated environments

Distraction as a measure of presence: using visual and tactile adjustable distraction as a measure to determine immersive presence of content in mediated environments

Rolf Nordahl · Dannie Korsgaard

Abstract To assess and improve the user experience in entertainment products, developers need results of evaluation methods, which in detail measure the relationship between the mediated content and the resulting media experience. This paper proposes a method applying adjustable distraction (AD) to determine presence as immersion (Lombard and Ditton in At the heart of it all: the concept of presence, Department of Broadcasting, Telecommunications and Mass Media, Temple University, 1997) at selectable events (approximated real-time). Two experiments were conducted to investigate its applicability in computer games and movies with respectively visual and tactile AD. The first experiment examined whether the experienced intensity in a survival-shooter game, measured through questionnaires, was proportional to results from the AD method. The intrusiveness of the AD method was also addressed in the experiment by comparing the immersive presence ratings in a between-groups design. The second experiment investigated whether heart rate measurements, intensity ratings and the results of the AD method with vibration as the distraction signal were proportional when test participants watched a movie clip. The outcome of the first experiment indicated that no significant intrusion is caused by the method. In addition, results showed no proportionality between the AD method and intensity ratings. However, as the AD measurements were supported by flow theory, it might be that the results from the AD method using visual distraction are giving a more comprehensive indication of presence as immersion (rather than just the intensity dimension). The second experiment revealed proportionality between the intensity ratings and the heart rate measurements, while the results from the tactile AD method were not proportional. We suspect that this was caused by the great variance found across the test participants' thresholds of perceivable vibration. Because of this, it is suggested that a thorough screening process is conducted pre-test if the AD method should apply vibration as the distracting stimulus.

Keywords Presence · Evaluation · Computer games · Visual · Tactile · Distraction · Measurement

1 Introduction

Development budgets for large-scale commercial entertainment products have over the years reached levels north of 20 million US dollars (Rosmarin 2006; Rose 2006). This entails that the development of this type of games or movies is of high financial risk. Applying content-related evaluation methods to determine whether a media product translates into a positive user experience might aid the developers in making correct decisions on the road to creating a well-selling product. Due to the need of such methods, the authors have initiated an investigation on whether presence evaluation methods customized from virtual reality (VR) research can be applied (Nordahl and Korsgaard 2008).

A significant field of study within VR research is *presence*, namely, the investigation of the degree to which a VR session makes a user feel like "being there" in the mediated virtual environment (Lombard and Ditton 1997). The concept of presence resembles phenomena known from studies in film- and computer games theory such as "the diegetic effect" (Tan 1995), "immersion" (Brown and Cairns 2004),

R. Nordahl (✉) · D. Korsgaard
Medialogy, Aalborg University Copenhagen,
Lautrupvang 15, 2750 Ballerup, Denmark
e-mail: rn@media.aau.dk

"the magic circle" (Adams and Rollings 2006) and "flow" (Csikszentmihalyi 1990), which have been recognized by many (Sweetser and Wyeth 2005; Adams and Rollings 2006; Ermi and Mäyrä 2005; Bordwell and Thompsen 2004) as essential parts of creating successful entertainment products. This is especially due to the belief that being present in an environment makes a person able to respond emotionally to events and characters contained within the mediated experience (Tan 1995).

Being present in a virtual environment is generally a very intangible concept incorporating a range of psychological structures. It can therefore be hard to measure and document (Baren and Ijsselsteijn 2004, p. 1). To overcome this problem, presence measurement methods have emerged, which apply different approaches in providing an indication of the amount of presence experienced. This paper attempts to adapt some of these existing presence measuring methods to give developers of entertainment products a tool to handle the intangibility of evaluating a mediated experience.

1.1 Limitations of using questionnaires in the development process

The common way of measuring presence is through post-test questionnaires (Baren and Ijsselsteijn 2004, p. 4). Analysis of the data derived from such questionnaires is generally used to determine presence as a product of the *entire* mediated experience. Fragments such as a clever scene transition, an unlocked game reward or a risky verbal line from a character, will therefore not be evaluated separately but as a part of the whole experience. Having a questionnaire returning a general presence indication as a result of a test session will not aid the game and film makers, as this indication does not give any hints as to what fragments worked as intended, and which should be improved during development. To examine fragments with questionnaires, the experiment could be divided into smaller sub-experiments meant only to focus on a single fragment. However, that would take a significant amount of time for each interesting fragment and it would not provide developers with information on how preceding parts of the entire experience impact on the experience of the fragment.

The ideal case, with respect to evaluation of fragments, would instead be a method which is capable of measuring presence real-time during the experiment. The results of the method could thereby be depicted as a continuous time/presence relationship curve (Fig. 1). This could enable developers to relate minimum and maximum readings on the curve with elements encountered in the test session at the particular time of the measurement. Through such a curve the developers would also be able to distinguish different levels of presence, which thereby would make it
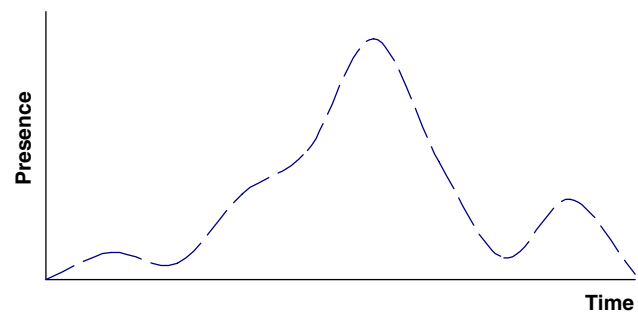

Fig. 1 The outcome of an ideal measuring method—example of a continuous presence *curve*

possible to track improvements in later tests by comparing the level of presence at the time where a certain part of the experience occurs.

Even though questionnaires are widely used in presence evaluation, we do not in general think they are designed for performing continuous measurements, due to the limitations of human memory. However, questionnaires might be used to approximate these measurements, but then they would have to contain questions regarding specific events or features in the movie, game or VR experience, which the developers consider to be important (e.g., unique selling points). As the participants have to evaluate retrospectively (post-test), these features or events have to be somehow memorable in order for the participants to remember the particular instance and how the experience was during that event.

1.2 Seeking inspiration in existing approximated continuous measuring methods

Several alternative presence measurement methods already exist. Even though not all of them produce moment-to-moment results displayed like a continuous curve (Fig. 1), they all seek to measure and log presence at points during mediation of the test material. One of them is the use of *psycho-physiological* measurements (measures of heartbeat, muscular responses, etc., as an indicator of presence). Applying physical reactions of the body as a measure of presence is justified by the theory that if one is present in a virtual environment, the body of the person will react to stimuli in the virtual environment in the same way as it would in the real world (Baren and Ijsselsteijn 2004, p. 44). Examples of these responses are increase/decrease in heart rate (Jang et al. 2002), changes in skin conductivity (Jang et al. 2002), eye tracking (Laarni et al. 2003) and changes in brain activity (Schlögl et al. 2002). A disadvantage of this type of measurement is generally that the equipment can pick up low signal-to-noise ratios. For instance, it has been seen that movements (electric activity) in the muscles around the head region (such as speaking or chewing) can

corrupt recordings of ongoing EEG measurements (Schlögl et al. 2002). Besides that, it is known that heart rate and skin conductivity correlate with emotions such as anxiety (Jang et al. 2002), which would normally only makes these measurements relevant to certain movie and game genres that seek to evoke and play on such emotional states.

Another suggested way of measuring presence is through *behavioral measurements*, namely observations of a person's behavior while being exposed to the test material. As with psycho-physical studies, these measurements rely on the idea that if a person feels present in an environment, the person will behave in the same manner in the virtual environment as in the real world, for instance by leaning in the counter-direction of a mediated movement (Freeman et al. 2000). A benefit of behavioral measurements is that behavior is often triggered spontaneously, without the test person having much control of it. However, a significant downside of this method is that the observer can misinterpret the behavior and that the test has to be specifically designed to a situation where a behavioral response can be provoked from the test persons, which is not necessarily the case for any feature or narrative in the environment which developers want to evaluate. Finally, it would probably require several observers to perform this kind of measurement, adding expenses to the overall budget.

The usage of *breaks in presence* (Slater and Steed 2000) also seems interesting as it can provide hints regarding what elements, in a mediated environment, are likely to induce breaks in presence during a test session (Steed et al. 2005). This way of measuring entails that the participant reports every time his or her attention is brought back to reality after the participant has been feeling present in the presented material. The method relies on the test participants' ability to judge whenever their attention has returned from feeling presence. A drawback of this method is that it is not sensitive to different levels of presence, and it does not take into account that a person can be partially focused on both the mediated and the real environment. Another issue is that it requires careful instruction of the test participants beforehand in order for them to be able to report when ever the feeling of presence is broken.

Besides breaks in presence other *attention-based* measuring methods have emerged in an attempt to measure presence continuously throughout an experiment. One variation is the introduction of a secondary task while having the participant performing in a VR experience. Darken et al. (1990) exposed test participants to a monitor displaying a movie with a detailed narrative, while the participant at the same time attended a VR experience. The assumption behind this method is that the amount of presence felt in a space corresponds to the amount of attention directed toward the mediated environment. By having the participants recalling events, which had occurred in either the movie or the VR experience, the attention directed toward each environment during the experiment could be measured. Thereby it could be determined at which times the participant had been more present in respectively the VR environment or the movie. Like breaks in presence, this method lets the developers know whether the participant has felt presence in their media product or not. However, with this method, it is still not possible to say at which level the participant has been feeling presence. Another point is that the questioning still takes place post-test (like questionnaires), and therefore the results still rely on the participants' ability to recall information.

A third attention-based method is the use of reaction time to a secondary task as a measure for presence (Baren and Ijsselsteijn 2004, p. 57). In a typical secondary-task reaction time (STRT) experiment, a test subject performs a task (e.g., plays a video game), and at some point, a signal (e.g., a tone) is used to alert the test subject (hence breaks presence). The time difference is then noted from when the signal was emitted to the time where the test subject responded. This time difference is then considered to be a measure of presence (Ijsselsteijn et al. 2000). The underlying assumption behind the STRT method is that as presence increases, more attention (and therefore more mental processing capacity) will be put in the performed task. If more capacity is put into performing the primary task, fewer resources remain for the secondary task (if a limited capacity is assumed), which, in the case of reacting on the signal, could result in a longer reaction time (Ijsselsteijn et al. 2000). Reaction time can on the other hand also be affected by a set of other factors. Reported factors are, for instance, muscular tension (Etnyre and Kinugasa 2002), age (Luchies et al. 2002; Philip et al. 2004; Der and Deary 2006), gender (Der and Deary 2006; Dane and Erzurumluoglu 2003), fatigue (Philip et al. 2004) and the breathing cycle (Buchsbaum and Callaway 1965). Experiments conducted with the purpose of testing the STRT method as an indication of presence have generally returned mixed results (Reeves et al. 1992, 1993; Klimmt et al. 2005). However, the method is interesting and a valuable contribution within the field of presence evaluation methods.

1.3 The adjustable distraction method

We think that an adaptation of one or more of these approaches might remove some of the weaknesses and add a valuable tool to the diversity of evaluation methods for entertainment media. The basic idea behind an adapted method is very well illustrated by the following example:

Imagine a conversation taking place between a man and a woman sitting on a couch. The television is running in the

background. Suddenly, the man catches something interesting on the television and redirects his attention from the conversation to the television. So he becomes more present in the world the television is offering and less present in the real world. When the woman realizes that he is not paying attention to the conversation, she tries to win back his attention. She does not feel like standing up and walking over to turn the television off; instead she attempts to say his name or make a rapid gesture in the hope of catching his attention. If this does not work, she might try to poke him or block his view to the television. Eventually, if none of the prior attempts worked, she would need to turn off the television, which without question will bring back his attention.

The point of the prior example is to demonstrate the assumption of the purposed method, namely the paradigm that *presence is as strong as the minimum amount of stimuli required to break it*. The woman in the example is obviously putting more and more effort into getting in contact with the man, and thereby she is increasing the amount of stimuli from reality directed toward him. It is then assumed, that if he is deeply present in a mediated experience it will require a stronger distraction to make him respond than if he only felt a superficial level of presence.

While thinking of the previous example the conceptualization of *presence as immersion* coined in "At the Heart of It All: The Concept of Presence" (Lombard and Ditton 1997) seemed to cover the concept well in this case. Presence as immersion is here described as "the degree to which a virtual environment submerges the perceptual system of the user" as well as "the degree to which inputs from the physical environment are shut out" and the extent to which the users feel "involved, absorbed, engaged, engrossed." Every one of these quotes is very descriptive about the situation of the man and his state of mind in the conversation–television example.

To form this idea into a concrete method, we took a starting point in the previously described attention-based secondary-task method (Darken et al. 1990). A number of test participants are here given a simple secondary task to perform, namely to react whenever they are exposed to a primitive stimulus. A "primitive" stimulus is not meant to be another mediated virtual experience such as it is the case in the secondary-task method. Instead, the stimulus should be regarded as a distraction, as in the STRT method, which immediately reminds the participant about the fact that he or she in reality is participating in an experiment and has to respond now. The stimulus used could be any perceivable signal (e.g., a visual, an audible or even a tactile signal). The important factor is that the signal can vary in strength, from undetectable to unavoidable. We have chosen to name the proposed method the adjustable distraction method (henceforth referred to as the AD method).

This method can be applied to most features or events in media products which might be a subject for evaluation. Improvements can also be tracked by applying the method to each iteration that the product goes through. The method also avoids any subjectivity issues, which for instance are found in cases where behavioral signs are misinterpreted by the observer. The AD method is comparable to breaks in presence as the participants are told to indicate (react) whenever presence is broken. However, in contrast to the breaks in presence method, the participants do not need a careful explanation of the presence concept in order to respond. The proposed method also has great resemblance to the STRT method in the sense that they are both measuring sustained attention. As such, the STRT method has all the advantages of the AD method (e.g., temporal measures, objective, sensitive to different levels, not retrospective). However, the AD method does not rely on time as a measure (but on the amount of distraction) and therefore all factors potentially leading to either increased or decreased reaction times (e.g., age, gender and recent muscular activity) can be disregarded in connection with the use of the AD method.

1.3.1 A mechanism for presenting adjusted stimulus

To make the AD method able to track levels of immersive presence, a mechanism for adjusting the strength of the distraction was needed. It has to be independent of time (e.g., the distraction should not gradually increase over time) as making it dependent of time would make all the time-related factors of the STRT a part of the AD method as well. We therefore propose the idea to adjust the strength of the distraction across test participants and thereby obtain a general indication instead of an indication from each individual. The algorithm works as the following.

A maximum (and minimum) strength of the distraction is given to the algorithm (e.g., strength 0 to strength 10). Half of this value should then be set as the initial distraction strength (e.g., strength 5) for every event, which is to be tested. Now an odd number (n) of test persons should play the game and when an event occurs, a distraction signal of the initial strength should be triggered with the participant either reacting (e.g., pressing a button) or not. The gathered "reacted-or-not" results would then be evaluated, for each individual event, by the computer. If the majority (over 50%) reacted upon the distraction signal of the initial strength, the distraction should decrease, for the next test round with half of the amount from the current strength to the minimum strength (e.g., strength 5 → strength 2.5). Likewise, if the majority did not react upon the distraction, the strength would increase with half of the amount from the current strength to the maximum

strength (e.g., strength 5 → strength 7.5). After the evaluation of the first test round, a new distraction strength for every event will be computed and a new round of n test persons can start playing the game with the appearance of the adapted distraction strengths. By always adjusting the distraction strength with either the upper or the lower half of the remaining range of untested strengths, eventually one strength will be left when enough test rounds have been evaluated. This distraction strength is assumed equal to the general amount of cognitive load (or maybe immersive presence) experienced at the particular event. The accuracy of this measure will improve as n tends toward infinity.

1.3.2 The intrusiveness of the method

In prior research, it has always been of great interest to keep participants in a stage of presence throughout an experiment, due to the risk of polluting the results through intrusion. Therefore, it might seem unwise to deliberately attempt to break it. However, it makes sense to break presence in order to measure it, when thinking of the analogy, that one has to disassemble something (which in some cases entails destroying it) to know how it is built. The intrusive effect of the method is of course of great concern, and therefore it must be a subject for testing, but we have not found any examples of why returning to the same stage of presence after a break, should not be any different than when, for example, a person is awoken by an alarm clock and then turn around and continue dreaming as nothing happened.

Optimally, each participant should only experience a single distraction which is strong enough to cause a break in presence. However, it would be interesting to test how intrusive additional breaks (repeated measures) would be, as more stimuli thresholds would make the method able to track how the feeling of presence evolves in the participants during the course of the test material, hence it would thereby approximate the output of an ideal measuring method.

2 Experiments overview

All of the considerations from the prior paragraphs can be lined up into one general hypothesis:

The AD method is capable of measuring presence as immersion at specific events during mediation of a virtual experience.

This hypothesis cannot be accepted based on a single experiment. However, a single experiment can indeed reject it. We therefore wish to initiate a series of experiments from which it is either possible to reject the hypothesis. If the experiments then show a supportive result it will contribute to the likelihood for the hypothesis to be true.

The remaining sections of this paper will describe the design, execution and implications of the first two of these experiments. Each of the two experiments will test the general hypothesis, but with an emphasis on different aspects which could be a potential problem in the execution of the AD method. A list of the two experiments is shown here:

Experiment 1: A test conducted in relation to a computer game in which a visual signal was used as the distracting stimulus.
Experiment 2: In this test, a movie clip was applied as test material while vibration was applied as the distraction.

In a section separating the descriptions, lessons learned and considerations regarding the results of the first experiment will be listed as well as how they could contribute to the optimization of the subsequent test.

3 Experiment 1: the visual distraction test

3.1 Addressed hypothesis

To test the "strength" of the general hypothesis, in this experiment, a set of sub-hypotheses was formulated to aid in the investigation of the AD method in this specific case where a game is the center of the test:

$H1_A$ A player's experience of intensity while playing a game is influenced to a degree which is proportional to the complexity of the presented challenge.

$H1_B$ A player's experience of intensity while playing a game is proportional to the minimum amount of visual distraction needed to attract the player's attention.

This last hypothesis is motivated by the fact that both the amount of mental processing and the degree of experienced intensity can act as indicators of presence as immersion.

$H1_C$ AD method measurements can be performed several times during a media experience without being significantly intrusive in relation to a participant's original experience of presence as immersion.

During the preparations for the experiment, we made a simplification in order to be able to measure presence as immersion at specific events (explained in a later paragraph). Due to this, simplification presence as immersion is

described in the hypotheses as the degree of "intensity." A confirmation of $H1_B$ would strengthen the idea that the AD method measures at least some part of presence as immersion. Examining $H1_C$ will inform about the level of "noise" which can be expected when applying the method. $H1_A$ is the basis of how we expect the level of presence to be during a computer game.

3.2 Experimental design

To study the intrusiveness ($H1_C$) of the AD method, a between-groups design with two experimental conditions was used. The independent variable was the AD method, which was applied on one of the playing groups, while the other group played without it. The depended variable was the extent to which the participants felt presence as immersion. By comparing intensity ratings given by both groups in post-test questionnaires, an insignificant variance in the means was expected in order for the AD method not to be intrusive.

The above-mentioned experimental design was also able to test $H1_B$. As part of the study, a section of the post-test questionnaire was designed to measure intensity at three events retrospectively. Then by comparing the results from the AD method (taken at the same three events) with ratings of intensity from the questionnaires, it could be concluded whether the results correlated or not. Normally, an experimental design will benefit from the use of a statistical model, but in the case of $H1_B$, it was not appropriate as the experiment encompassed two very distinct measurement methods, from which the results will be of different scales. Scaling the results to be within the same span was not considered an option, as the minimum and maximum values feed to the algorithm controlling the adjusted distractions are estimates given by the experimenter. Instead, the comparison was based upon the relationship between the mean values of the results collected for each event.

3.3 Test material

A rough game prototype, centered on a first-person shooter survival gameplay, was created for the purpose of this test. The prototype was meant to represent some of the early drafts and stages that a game developed through a rapid prototyping methodology goes through. This entails that the game contained all the basic gameplay elements but in a crude an unpolished state. Settled in a virtual house the player was here given the main task to survive for 5 min while fighting off an increasing number of enemies (in this case zombies).

3.4 Data collection

Besides the AD method, post-test paper and pencil questionnaires were used for data collection.

The part of the questionnaires examining $H1_C$ were developed according to the measurement guidelines provided by the international society for presence research (ISPR 2000), which state that terms such as intense, fun, competitive, addictive and exciting can be used in the formulation of questionnaires treating immersive presence. This lead to the following seven questions accompanied by Likert scales (Likert 1932) with seven rating levels to enable the participants to grade each case:

> To what extent did you forget about the real surroundings while playing the game?
> How much fun did you have playing the game?
> How exciting was the game?
> How enjoyable were the tasks and challenges presented to you in the game?
> How intense was the game?
> How hard did you try to achieve the goal(s) presented to you?
> How much would you like to play the game again?

Each of these formulations is meant to cover some part of the concept of presence as immersion, and by asking participants to rate all seven questions, we hoped to get a useful indication of this multidimensional structure.

Due to the multidimensional nature of presence, it is usually described through several words which each represent some part of the concept (Baren and Ijsselsteijn 2004, p. 4). As it was assessed that it would be too overwhelming and monotonous for the test participants to answer the same three event-questions with different ISPR-terms (e.g., using seven terms to cover presence would result in $3 \times 7 = 21$ similar questions), it was decided to focus on a single descriptive word for the questions. The three events to be rated were selected based on the game prototype. The game was largely about stressing the player by exposing him or her to lots of attacking enemies. A set of situations that could arise in connection with this could be to name a few, events with many enemies near the player combined with little ammunition, which were assumed to create a more *intense* experience than events with just many enemies (because of the limited means available to deal with the enemies). As a result, the formulations were as follows:

(1) *How intense was the game when you had <u>low ammunition</u> and were attacked by <u>many</u> zombies?*
(2) *How intense was the game when you had <u>medium ammunition</u> and were attacked by <u>moderate amount of</u> zombies?*
(3) *How intense was the game when you had <u>high ammunition</u> and were attacked by <u>few</u> zombies?*

Each of these questions also applied Likert scales with seven rating levels which enabled the participants to grade

Table 1 Displaying the set of conditions describing when the three events will appear

	1. Low ammo and many zombies	2. Medium ammo and moderate zombies	3. High ammo and few zombies
Zombies in the room	More than 5	From 3 to 4	From 1 to 2
Amount of ammunition	From 0 (empty) to 6 shells	From 7 to 14 shells	From 15 to 22 (full) shells

each case. It was expected that the level of intensity experienced would be descending from event 1 to 3. For example, moments with low ammunition and many enemies nearby was assumed to be more intense (and maybe more immersive) than moments with few zombies and lots of ammunition.

3.5 Implementation of a visual adjustable distraction

The three events to trigger the distraction were specified in the game code through "if" conditions. The amount of zombies encountered by the player was here simplified as the amount of zombies in the same room as the player. Table 1 lists the conditions that should be fulfilled in the game before an event was registered as occurring.

The assumption was that the player during a play session would treat the area around the center of the screen as the main point of visual attention, while the outermost areas of the screen would stay in the periphery of sight. As described in Lombard and Ditton (1997), the conceptualization of presence as immersion should entail that the test participant would shut out the world, including perception of the visual periphery. If the distracting stimulus was then presented in the outermost areas one should be able to tell when a break in presence occurs. Placing a black border (a neutral zone) around the game visuals makes up an edge, which separates the screen into two regions (Fig. 2). The gradually changing visual stimulus was made by altering the size of a gray circle, appearing within the black border on the screen. The larger the circle, the stronger the distraction was assumed to be. The appearance of the circle or dot of a certain size should then either go unnoticed or work as a distraction and break the sense of presence and make the participant react by pressing the right button on the mouse. To hinder that the player would start to anticipate, the dot at a specific location random placement of the dot somewhere within the neutral zone was used.

The maximum diameter of the dot was set to the width of the black border such that the dot size could vary from 0 to 128 pixels. Then, whenever an event would occur, the participant was exposed to a distracting stimulus of certain strength within this span. The participant was then given 2-s to press the button before the stimulus would disappear and the event was marked as "not perceived."

Fig. 2 Screenshot of the game prototype where the screen area is separated into two regions. One region contained visuals presented in relation to the secondary task, while the other region displayed the content of the game

3.6 Equipment

The setup consisted of a computer capable of running the game prototype with and without the implementation of the binary search algorithm. A 15.4″ monitor (running the game full screen in a $1,024 \times 768$ resolution) was connected to the computer as well as mouse and keyboard for the participants to interact with the game and respond to the distracting stimulus.

3.7 Procedure

The test was conducted in the facilities of Medialogy, Aalborg University Copenhagen. A total of 66 people, between ages 20–30 years, participated and all were full-time students. All participants played first-person shooters on a daily basis and were experienced, and none of them disliked violence in games. Thirty-three participants were assigned to test the game using the AD measuring method (AD group), while the other 33 participants would test the game while no measurements were taken (control group). It was in general very difficult to find female participants who fulfilled the criteria for participation in the experiment (especially fulfilling the criteria of being experienced in playing first-person shooter games). Only four women could be recruited for each group, making the male participants the decisive gender. As the experiment was initiated to give an initial impression of the usefulness of the AD method (and not as a study in gender related

Table 2 The listed mean, standard deviation and the probabilities of significance for each of the seven presence as immersion questions in the questionnaires

Parameters	Control group	Adapted group	Probability
The degree to which the real surroundings was forgotten	4.36 ± 1.49	4.33 ± 1.81	$p < 0.946$
The degree of fun	4.87 ± 1.05	4.36 ± 1.14	$p < 0.067$
The degree of excitement	4.57 ± 1.06	4.12 ± 1.59	$p < 0.201$
The degree of enjoyable tasks	4.63 ± 1.05	4.03 ± 1.28	$p < 0.086$
The degree of game intensity	4.36 ± 1.24	4.09 ± 1.48	$p < 0.45$
The degree of effort put into achieving the goal(s)	4.91 ± 1.31	5.12 ± 1.59	$p < 0.579$
The degree of addictiveness	4.68 ± 1.18	4.72 ± 1.55	$p < 0.924$

issues), the fact that both groups had the same gender ratio was considered to be sufficient.

Before playing the game, all participants were told about the objectives in the game and what the controls were. The distraction group was given additional instructions to press a mouse button when the dot would appear. Participants of the distraction group played the game at intervals of five persons per test round ($n = 5$). After the play sessions, all participants were asked to fill out the questionnaire.

3.8 Results

The means and the standard deviations of the over-all presence questions is displayed in Table 2 for each group. The last column lists the probabilities found through a paired t-test assuming equal variances between the means from both groups. All the means are based on a scale where 1 was regarded as a low degree and 7 was regarded as a high degree.

The means and standard deviations from the event-based part of the questionnaire are listed in Table 3. Like the seven prior questions, the event-based questions were likewise based upon ratings where 1 was regarded as a low intensity and 7 was regarded as a high intensity.

A two-way ANOVA analysis showed insignificant variance between groups, $F(1,192) = 2.32$, $p < 0.128$, and an expected significant variance between events, $F(2,192) = 12.93$, $p < 5.83\text{E-}06$, with no significant interaction, $F(2,192) = 0.45$, $p < 0.638$. The results from the adjustable distraction method are shown in Table 4 with respect to each event.

Assuming the adapted method is valid as a measurement of immersive presence, the size of the dot should correspond to the general strength of distraction needed to break a player's concentration during the given event. "Times evaluated" differs between events because some events occurred less often than others. As a limited amount of test participants were at disposal, the binary search algorithm did not manage to reach a single dot size value for all of the events. Therefore, another parameter must be listed with

Table 3 The mean and standard deviation of the intensity ratings with respect to each event

Parameters	Control group	Adapted group
1. Intensity, low ammo and many zombies	5.38 ± 1.29	5.31 ± 1.56
2. Intensity, medium ammo and moderate zombies	4.26 ± 1.36	3.83 ± 1.19
3. Intensity, high ammo and few zombies	3.36 ± 1.76	2.96 ± 1.49

the results. "Pixel uncertainty" refers to the maximum amount of pixels (px), which the dot size can vary at the arrived point in the search algorithm. The last column "Visual relationship" gives a visual indication of the relationship in size between measurements of each event.

3.9 Discussion of the visual distraction test

The t-test scores from the seven presence as immersion questions show a relatively low p value related to the degree of fun and the degree to which the tasks were enjoyable. This could be an effect of the AD method. However, none of the group means are significantly different from each other (and the difference in means could therefore also be an outcome of the standard error). This result supports $H1_C$ and suggests that the AD method was not as intrusive as it was feared to be.

Results from the event-questions in the questionnaire (Table 3) show that the participants from both groups ranked the events as expected ($H1_A$). However, the results from the AD method (Table 4) show a different pattern. Event 1 was expected ($H1_B$) to have the highest value, but is here assigned the lowest. At a first glance, these results might seem very irrational. However, after some considerations it appears that the results could in fact be supported by the thoughts of the challenge/skills relationship in flow theory (Csikszentmihalyi 1990). According to this theory, challenges encountered while performing an activity should be balanced in relation to the player's

Table 4 Results returned from the adjustable distraction method accompanied by a visual representation

Event	Final diameter (px)	Times evaluated	Remaining uncertainty (px)	Visual relationship
1. Low ammo, many zombies	28	4	±3	
2. Medium ammo, moderate zombies	94	5	±1	
3. High ammo, few zombies	58	5	±1	

skills if the experience is to generate a feeling of flow (a state of mind with features similar to those of presence as immersion). Considering the three chosen events as three challenges of distinct difficulty, it makes sense that event 2 (medium ammo/moderate zombies) required the strongest distraction for players to respond. If games are either too easy (e.g., high ammo/few zombies) or too hard (e.g., low ammo/many zombies), they can become either boring or overwhelming, which might lead the player to direct his or her attention somewhere else. Since presence as immersion has some similarity with flow, for instance, flow can only occur while a person is feeling a deep effortless involvement that removes the awareness of thoughts and other stimuli presented by real life (Csikszentmihalyi 1990, p. 49), the results from the AD method might in this case be giving a more "complete" indication of presence as immersion at the specific events rather than just the single intensity dimension addressed by the questionnaires. It also can be questioned whether the answers to the event-questions (Table 3) actually are the outcome of the same logic thinking which made up our expectation ($H1_A$) in the first place. Some of the events had probably occurred several times during a single play session (e.g., high ammunition and few zombies) and maybe events were taking place over a longer period of time (e.g., several minutes). This must have made it difficult for the participants to assign memorial moments to each of the question categories, and therefore we speculate that they might have relied on logic instead when rating the three event-questions. To conclude, the results either confirms $H1_A$ and rejects the $H1_B$ or in the worst case are inconclusive in both the matter.

While conducting the experiment, we observed a range of problems we had not initially thought of. One issue was that the total impression of illumination coming from a screen changes according to the content displayed. For instance, if a player is playing a first-person perspective game, as in the first experiment, and walks into darker areas the perceived luminance of the screen will decrease. This damping might have emphasized the visual stimulus in a way that made it appear stronger than intended. Another problem was the effect of having the black frame surrounding the game visuals, as previous presence studies have found that applying smaller displays will decrease the feeling of presence (Lombard et al. 2000). Lastly, the random position of the dot in the black frame might potentially make the dot appear right next to the enemy whom the player are currently looking at. The user's point of attention on the screen should ideally be at the same distance from the dot each time it appears. To filter out some of these uncertainties, we instead used vibration as the distracting stimulus in the next experiment as this type of stimulus operates outside the audiovisual space, which most media occupy.

4 Experiment 2: the tactile distraction test

4.1 Addressed hypothesis

The first experiment implied that the AD method does not correlate with the amount of intensity in a computer game. In the forming of the hypotheses for the second experiment, we asked ourselves what the AD method instead was

indicating. The nature of questionnaires have issues which do not make them suitable for measuring real-time effects and therefore another measuring method might aid in the investigation of the AD method. Psycho-physiological heart rate (HR) measurements have earlier produced promising results in connection with scary (Jang et al. 2002) or violent (Cameron and Ballard 2002) test material. It is therefore interesting to examine whether the results from the AD method would correlate with the results obtained by HR measurements. However, this required that the test was done based on material of a relevant genre or theme. The sub-hypotheses for the second experiment are as follows:

$H2_A$ A person's heart rate recordings during the experience of a conventional thriller movie are proportional to the suspense build by the narrative.

$H2_B$ A person's perceived intensity of events in a conventional thriller movie is proportional to the suspense build by the narrative.

$H2_C$ A person's heart rate during a conventional thriller movie experience is proportional to BOTH a person's perceived intensity of events in a conventional thriller movie AND the minimum amount of tactile distraction needed to attract the person's attention (because the amount of mental processing and the changes in heart rate and perceived intensity due to suspense in fact all are indicators of presence as immersion).

$H2_B$ is included in the experimental hypotheses as we speculate that the non-linear nature of games made it hard for the participants to answer the three intensity ratings in the questionnaires. As a movie clip is applied as test material in this experiment, we wished to examine whether the intensity rating would correlate with the results from the AD method when applied to a media with a linear narrative structure. $H2_C$ will examine whether all of the three included measurement methods are monitoring the same phenomenon.

4.2 Experimental setup

The same methodology as in the first experiment was applied. Each method was to collect repeated measurements during three selected events in a movie. It was then expected that the results from each method would carry the same characteristic pattern as if the measuring methods in fact were measuring the same phenomenon ($H2_C$). As with the first experiment, a comparison between the obtained measurements is only possible based on the relationship between the means of each event as the scales are different. Scaling the results to the same span is not considered to be justifiable. If an unexpected result would occur in the

examination of $H2_C$ it might be necessary to re-evaluate the validity of $H2_A$ and $H2_B$.

4.3 Test material

We speculate that the developed questionnaires from the first test caused problems when test subjects had to rate events as the questions referred to reoccurring events (which computer games are full of) and events which appeal more to the use of logic thinking rather than answers based upon the memories from the actual experience. To apply questionnaires in comparison with the AD method, the computer game used as test material in the first experiment was instead replaced by a movie clip.

Regarding choice of the test material, the movie clip had to contain a storyline with events that had a reasonable chance of inducing distinct levels of presence, as the characteristic pattern found in the results across other measuring methods and the AD method would seem more convincing than if a somewhat monotone pattern was found. Another thing to look out for was the possibility that persons appearing in the movie clip (e.g., a famous actor/ actress) might have contributed to other media productions experienced by the participants. Prejudices or certain views toward a particular person may affect how the movie clip was experienced, which was undesirable. This makes it so that most Hollywood movies, reusing known actors, were inappropriate for the use as test material.

By performing a range of small viewer tests on individuals of the target group, the movie "El Laberinto del Fauno" (del Toro 2006) seemed to be a good candidate in accordance to the list of demands. With its Spanish list of actors, the movie has a relatively unknown cast (seen from a Dane's point of view), which should minimize the occurrences of prejudices among the test participants. During the viewer's tests, Chap. 9 (and some of Chap. 10), were shown from the Scandinavian DVD version. To quickly summarize the action, a little girl has found a secret door leading to a room where a monster-like pale man is sleeping. While the girl is occupied the pale man wakes up and starts to hunt her back to the door opening. Even though the door shuts in front of her she manages to escape, before the pale man reaches her, through another door in the ceiling. According to the descriptions given by the audience, this scene was perceived to almost constantly increase in intensity and peak at the sequence where the key figure attempts to escape. Some of the individuals also referred to the pale man as "suddenly appearing comical" during a period in the scene where it is revealed that the eyes of the pale man are located in the palms of his hands. This change in perception from scary to comical might lower the immersive presence felt by the audience due to the inconsistency of a comic element in a very consistently

scary setting. Based upon the statements from the viewer's tests, the following events were selected for this experiment:

Event 1: Sequence where the pale man suddenly appears comical; medium level of presence expected (appears 4 min 26 s after the Chap. 9 mark)

Event 2: Sequence of the escape from the pale man; high level of presence expected (appears 6 min 5 s after the Chap. 9 mark)

Event 3: Beginning of the subsequent chapter; low level of presence expected (appears 6 min 50 s after the Chap. 9 mark)

The sequence chosen for event 3 was the beginning of the next scene which by all means is out of context in relation to the two other events. As the participant had no reference for knowing any of the characters or the setting of the scene, a low level of presence was expected here.

4.4 Data collection

Along with AD method measurements and questionnaires, heart rates were recorded with a Biopac Science Lab MP40 device.

The questionnaire had to contain new formulations of the event-based questions, adapted to fit the selected events. In general, it seemed to be the selection of logic events that resulted in a potential unreliable rating in the first test. It was therefore decided to keep "intensity" as the describing word for presence as immersion in the questions. The following three formulations were therefore asked in the questionnaires:

(1) *How intense was the sequence where the pale man puts his hands (with his attached eyes) up in front of his face to look at the girl for the first time?*

(2) *How intense was the sequence where the door closes in front of the girl, preventing her from escaping?*

(3) *How intense was the sequence where the group of people is walking through the woods?*

These questions were built around a 5-optioned Likert-scale with a rating where 1 was regarded low intensity and 5 was regarded high intensity.

4.5 Implementation of vibration as the adjustable distraction

The implementation of the AD method was, in this experiment, strongly inspired by the setup of the vibration threshold experiments conducted by Verrillo (1962). To be able to expose test participants to different strengths of a vibration stimulus, a small pc speaker was used. Speakers are basically an electromagnet that reacts

according to the amount of voltage put through it. When voltage is mapped according to a sinusoid and send to the speaker, the electromagnet will alternately start to attract and push away the membrane creating vibrations. The characteristics of the vibration can be read out of the sinusoid signal sent to the speaker. Two parameters are of significance here, namely, the frequency and the amplitude (the displacement of the speaker membrane) and changing these quantities will therefore alter the properties of the vibration. As the AD method only allows for regulation of a single parameter to vary the strength of the stimulus, one of these parameters had to be kept fixed while the other could vary. Observing that frequency thresholds, in Verrillo (1962), displayed as a *U*-shape curve, favored (due to the non-injective nature of the frequency thresholds) that the frequency was kept fixed while the displacement/voltage parameter was used as the variable. The frequency of the speaker was to be kept constant at 250 Hz because the skin is most sensitivity to vibrations of this frequency (Verrillo 1962).

4.6 Equipment

The setup consisted of a computer connected to the pc speaker (and an amplifier), through the sound card's output port. A program, running at the computer, maintained the frequency at 250 Hz and adjusting the displacement through an implementation of the adjustable distraction algorithm. To carry on the vibration, the head of a 5 mm × 50 mm stainless steel A4 bolt (non-magnetic material) was attached to the center of the speaker membrane. The speaker was then placed onto a jack under a table in a way that the pointy bolt had free passage through a hole (with diameter of approximate 12 mm). By adjusting the jack, the bolt could then be positioned with precision in relation to the upper side of the table top (Fig. 3). Besides the speaker, a small USB numerical pad was connected to the computer. The "Enter" key was here marked with a small label to emphasize its significance above the rest of the buttons. The numerical pad was then placed on the table top a small distance to the right of the hole with the bolt.

The table, speaker, numerical pad and a chair, for the participants to sit on, was located in a sound proof room. However, the computer was placed in an adjacent room to avoid that the appearance or the noise of it would interfere with the experiment. In front of the table, a 32″ LCD screen was used to display the movie clip. To provide the participants with the sound of the test material, a set of muffled headphones was connected to the screen. Besides mediation of sound, the used headphones also filtered out any remaining sound coming from the vibrations of the pc speaker.

Fig. 3 The setup. *1* The stainless steel bolt attached to the pc speaker, *2* a cord connecting an ohm meter to the bolt, *3* a digital ruler, *4* the jack, *5* a mirror used for reading the digital ruler while operating the setup

4.7 Procedure

The experiment was conducted in laboratory facilities at the campus of Aalborg University in Copenhagen. The participants had not seen the particular movie clip prior to the experiment and did not suffer from any heart disorders (as far as they knew). As large parts of the clip are without any speech, it was decided to show it without subtitles to give all the same conditions independent of language skills. It was therefore ensured that none of the participants knew Spanish (the spoken language in the movie). In addition, the participants were kept indoors for at least and hour before the experiment to ensure normalized skin temperatures as low temperatures can have an effect on vibration perception (Verrillo and Bolanowski 1986). Twenty-three participants of both genders between ages 20–30 and all full-time students were part of this test. Each participant went through the procedure described in the following.

The participant was seated in the chair and asked to place the left hand at the numerical pad (with their index finger at the "return" button) and to place the right hand on the surface of the table top just above the bolt-hole. The right hand was placed in such a way that the bolt would hit around the center of the first metacarpal of the thumb (the large pad in the palm just below the thumb). This contact area was chosen as it was expected to be the area on the
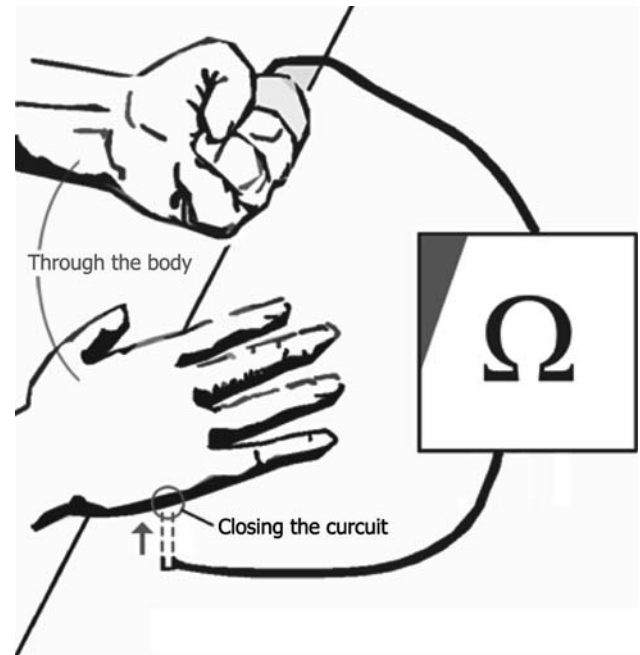

Fig. 4 The circuit that was used to ensure that the same pressure was applied on the skin of the participants

hand less likely to move during the experiment. The participants' hands was to remain unclamped during the entire experiment as it previously have been shown no to have any significance in relation to the precision of the measured thresholds (Verrillo 1962). Instead, they were told to keep their right hand still. When both hands were in place the participant would be asked, for a short period, to hold one end of an Ohm meter in the left hand while the other end was being attached to the bolt at the speaker unit. Now the jack was used to make contact between the two ends of the circuit by gently decreasing the distance between the bolt and the skin of the participant (Fig. 4). As soon as the Ohm meter displayed a reading, the bolt would just barely touch the skin. To ensure full contact, the bolt was then raised 1 mm. This process is necessary as the level to which the bolt is pressed toward the skin has an effect on the perception of vibrations (Verrillo 1962).

The next phase of the experiment was dedicated to an investigation of the vibration threshold when the participant was unaffected by the test material. In this process, the participant was asked to wear the headphones while he or she was being exposed to different levels of vibration. The adjustable distraction search algorithm was used here, but only on the single participant ($n = 1$) to find the minimum level at which the vibrations were perceived. Instructions were given that the Enter button on the numerical pad should be pressed if vibrations on the hand were perceived. During this phase, the headphones emitted white noise to make sure that the vibrations were felt and not heard. As white noise (as well as silence) is assumed not to contain

any immersive content it served the additional purpose of removing yet another uncertainty. Since the sound track of the movie clip replaced the white noise in the headphones in the subsequent phase of the experiment, participants would be exposed to vibrations both at the ear level (vibrations of the sound) and at their right hand (vibrations from the bolt). Therefore, white noise was used during the initial phase to create similar ear vibrations as it is not known whether these vibrations would influence the perception threshold of the vibrations sensed in the hand. Likewise, noise was displayed on the LCD screen to ensure that light coming from the screen in both phases of the experiment would not have a significant effect on the vibration results. When the threshold was identified and registered, the actual experiment could begin. The test participant was then equipped with electrodes (at both ankles and the right wrist) to be able to measure changes in heart rate (HR) during the playback of the movie clip. The participant was then reminded that the button on the numerical pad should be pressed if vibrations in the hand were noticed during the movie experience. The playback of the movie clip and the program which managed the occurrence of events was then started simultaneously to ensure synchronization. After the movie clip the participants were asked to fill out the post-test questionnaires.

4.8 Results

As the relationship between the event-means obtained through each method is of importance, the results will be listed as graphs. This is helpful as patterns and scale relationships are quickly revealed in a graphical representation. The results obtained through the questionnaires are given in this section. Both mean and standard deviations are given in Fig. 5 with respect to each event.

The HR measurements were treated by reading the rate in beats per minute (BPM) at the time where the particular event was recorded. The means and the standard deviations are displayed in Fig. 6 with respect to each event.

The first part of the experiment, determining the threshold for perceivable vibration of each test participant, returned a mean threshold of 275.2 millivolts (mV) amplitude from peak-to-peak (p–p) of the sinusoid signal provided to the speaker. The standard deviation of the thresholds was rather high at 177.5 mV p–p. Results from the adjustable distraction search algorithm should therefore be seen in accordance with each individual threshold as displayed in Table 5. The rows (each participant) in Table 5 are listed in the order they were tested. As each evaluation round was set to five persons ($n = 5$), either no reaction or a reaction three times (the majority would already be reached) at the same vibrations strength would
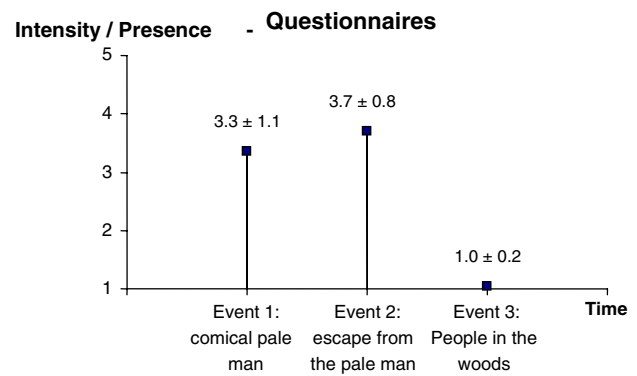

Fig. 5 Mean and standard deviations from the event-based questions in the questionnaire. Answers were given on a 5-optioned scale where five was more intensity and one was less intensity
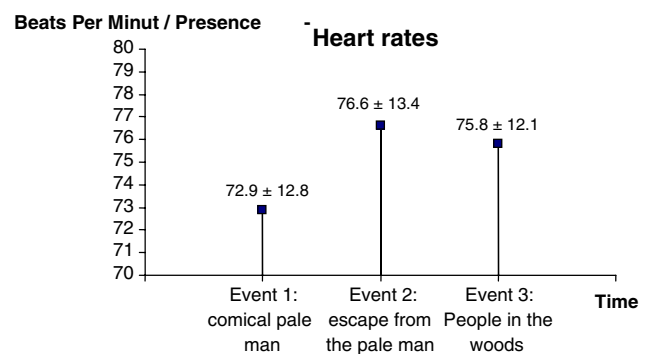

Fig. 6 The mean and standard deviation of participant's heart rates registered at the time of each event

results in the adjustment to a new evaluated stimulus strength.

The event-columns in Table 5 have been marked with colors to illustrate whether the test participant responded on the stimulus or not. White cells show that the participant reacted on the stimulus, while the gray cells are cases where the participant did not press the button after being presented to the stimulus. The black cells are interesting. These are instances where the participant would have reacted to the stimulus if the participant had only been exposed to noise (through the screen and the headphones), but did not do so when exposed to the movie clip.

As each evaluation round was set to five persons ($n = 5$), either no reaction or a reaction three times at the same vibrations strength would results in the adjustment to a new evaluated stimulus strength (due to the fact that the majority would already have been found even though five participants have not been tested yet). The values at the doubled underscored columns at the bottom of the table are the results returned by the adjustable distraction search algorithm with respect to each event. At all the events, 215 mV peak-to-peak was returned.

Table 5 Results from the AD method displaying perceivable vibration thresholds for each participant and strength of the vibration at each event

Participant Index no.	Threshold mV p-p	Event 1 mV p-p	Event 2 mV p-p	Event 3 mV p-p
1	462	2140	2140	2140
2	216	2140	2140	2140
3	247	2140	2140	2140
4	191	1094	2140	2140
5	206	1094	1094	1094
6	247	1094	1094	1094
7	194	566	1094	1094
8	410	566	1094	566
9	562	566	566	566
10	97	566	566	566
11	318	315	566	315
12	309	315	315	315
13	138	315	315	315
14	488	188	315	188
15	185	188	315	188
16	162	188	315	188
17	122	188	188	188
18	862	188	188	188
19	215	247	188	247
20	203	247	188	247
21	144	247	247	247
22	156	247	247	247
23	197	247	247	215
		215	215	215

The stippled lines mark the end of each evaluation round. The white cells are strengths with a reaction, gray cells are strengths without a reaction but below the perceivable vibration threshold and the black cells are strengths without reactions but above the perceivable vibration threshold

4.9 Discussion of the tactile distraction test

Prior to the experiment, it was expected that the results from both questionnaires and heart rate measurements would be similar to the results obtained through the AD method. The questionnaires (Fig. 5) and the HR measurements tend to correlate in their measures at event 1 and 2 (event 1 was rated lower than event 2). However, the methods do not return the same indication of how event 3

should be rated according to the two other events. The questionnaires returned a very low score where the HR measurements indicated medium intensity. A speculation, which might explain why HR results differ in the last event, could be that a transition from one scene to another occurred in the movie clip between event 2 and 3 where new characters and a new location were shown to the participants without proper introduction. The incoherence of the latter scene could have left the participants in a state where they might expect an imminent closure to the experiment. This new intense "real-life" situation could have raised the test participants' heart rate to a relatively high level while they still rated the third event in the movie clip with a low intensity score. With this as a very likely cause, the $H2_A$ and $H2_B$ could be accepted.

The results from the AD method show a very distinct pattern from the questionnaires and the HR measurements. Here, the same stimulus strength was returned for all the events. This clearly underlines that the AD method *in its current form* does not measure the same phenomenon as the two other measuring methods and that $H2_C$ therefore can be discarded.

While discarding the hypothesis, the authors want to emphasize the words "in its current form." It was unexpected to find this great variance in the recorded vibration thresholds, due to the consistent findings in Verrillo (1962). This entails that the vibration strength presented at an event probably is perceived very differently according to the threshold of the test participant. For instance, it could be reasoned that a vibration displacement of 300 mV p–p would be perceived stronger by a person with a threshold of 150 mV p–p, than a person with threshold 290 mV p–p. The adjustable distraction algorithm, originally designed in relation to a visual stimulus, showed to be vulnerable to this variance especially when using a low number of test subjects per evaluation round (n). For instance, the 188-strength-evaluation round in Table 5 is entirely determined by the non-responses made by test participants, which had a higher threshold 188 mV p–p. In experiment 1 where a visual stimulus was applied, the perceivable thresholds were not this big an issue as the screen is build from discrete pixels. Here, all test participants with a normal sight probably had a perceivable threshold of a circle of 1 pixel. To eliminate the influence of significant variation in perception thresholds of the used distraction signal, a thorough screening process should be conducted pre-test.

Finally, it is worth mentioning the black cells in the tables from the AD method. These cells represent all the cases where test participants did not react on the presented stimulus even though they were expected to do so. This indicates that the movie clip is the decisive factor in whether a response was triggered or not in over 26% (6/23) of the test participants. We suspect that the underlying

reason behind the missing responses is that the movie clip did induce more immersive presence than the mediation of noise. It therefore demanded a stronger distraction in order to break. This indeed supports the paradigm of the AD method, namely that presence is as strong as the minimum amount of stimuli required to break it.

5 Conclusion

The adjustable distraction (AD) method has the potential to objectively measure immersive presence at chosen events during any media experience. The method is based on the tracking of responses to stimuli of a variable strength, which are adjusted in order to find the minimum stimuli strength at which presence as immersion can be broken. Its usefulness was investigated in two cases, each applying a distinct stimulus (visual/tactile) and a different media (computer game/movie).

The results returned by experiment 1 made it clear that the AD method applying a visual distraction could perform measurements several times during a test session without being significantly intrusive in relation to the test participants' original experience (confirmation of $H1_C$). The results of the first experiment also showed that the AD method is not proportional to the degree of experienced (or at least the rated) intensity in a computer game (rejection of $H1_B$). However, the results also suggest that the measurements from the AD method using visual distraction might be giving more comprehensive indication of presence as immersion (rather than just the intensity dimension) as the results were supported by flow theory.

Experiment 2 produced results confirming $H2_A$ and $H2_B$, namely that the heart rate and the perceived intensity of the movie experience was proportional to the suspense build by the narrative. However, the results investigating $H2_C$ indicated the AD method cannot be used with tactile distraction if a proper screening process (to level out threshold variances) is not included in the experiment.

Therefore, the next logical step for future experiments would be to incorporate a screening process when applying tactile distraction as the adjustable stimulus. Furthermore, the implications of the experiments are conditional on the possibility that results from this study can be transferred to a population of an equal gender ratio. Considering that both executed experiments were conducted on groups with a majority of male participants makes it evident that future work is needed, which encompasses gender balanced groups.

6 Future perspectives

In the future, a range of aspects regarding the AD method are subjects for experiments. First and foremost research could

be conducted on how to further develop the AD method to encompass another mechanism to present adjustable distractions, which account for the variance in perception thresholds. This initiative would in theory (if it is possible to develop) entail that all test subjects in each evaluation group are exposed to a stimulus which is perceived to be of equal strength. In other words: even though each participant in an evaluation group is exposed to vibrations which differ in amplitude, all participants would (due to differences in, for e.g., fat percentage, amount of hair) perceive the particular vibrations as being of the same strength.

As the AD method showed promising results with the use of a visual signal as the distraction more extensive research and development of the method could be interesting. For instance, eye-tracking techniques could be applied to make sure that the visual dot always is at the same distance from the gaze of the participant. Image processing techniques could also be utilized to regulate the visual stimulus according to, e.g., the mean brightness of the screen in order to ensure that the strength of the stimulus is perceived equally independent of the light and colors of the virtual environment.

As the two experiments have been very concerned with the evaluation of the content of media products, future studies could attempt to evaluate the relevance of the AD method in the context of media form (e.g., screen size, surround sound vs. stereo). As form has been reported (Lombard et al. 2000) to influence presence, the AD method might also have potential here.

References

Adams E, Rollings A (2006) Fundamentals of game design, 1st edn. Pearson Prentice Hall, Englewood Cliffs, NJ

Baren J, Ijsselsteijn W (2004) Measuring presence: a guide to current measurement approaches. Available via: http://www.informatik. umu.se/~jwworth/PresenceMeasurement.pdf. Accessed 30 Jan 2009

Bordwell D, Thompsen K (2004) Film art: an introduction, 7th edn. McGraw Hill Higher Education, New York

Brown E, Cairns P (2004) A grounded investigation of game immersion. CHI '04 extended abstracts on Human factors in computing systems. ACM, New York

Buchsbaum M, Callaway E (1965) Influence of respiratory cycle on simple RT. Percept Mot Skills 20:961–966

Cameron P, Ballard M (2002) High versus low aggressive priming during video-game training: effects on violent action during game play, hostility, heart rate, and blood pressure. J Appl Soc Psychol 32(12):2458–2474

Csikszentmihalyi M (1990) Flow: the psychology of optimal experience. Harper Perennial, New York

Darken R et al (1999) Quantitative measures of presence in virtual environments: the roles of attention and spatial comprehension. Cyberpsychol Behav 2(4):337–347

Dane S, Erzurumluoglu A (2003) Sex and handedness differences in eye-hand visual reaction times in handball players. Int J Neurosci 113(7):923–929

del Toro G (2006) The Spanish movie "El laberinto del fauno". Official site available via: http://www.panslabyrinth.com. Accessed 30 Jan 2009

Der G, Deary I (2006) Age and sex differences in reaction time in adulthood: results from the United Kingdom health and lifestyle survey. Psychol Aging 21(1):62–73

Ermi L, Mäyrä F (2005) Fundamental components of the gameplay experience: analysing immersion. Proceedings of the DiGRA conference 2005

Etnyre B, Kinugasa T (2002) Postcontraction influences on reaction time (motor control and learning). Res Q Exerc Sport 73(3):271–282

Freeman J et al (2000) Using behavioral realism to estimate presence: a study of the utility of postural responses to motion stimuli. Presence 9(2):149–164

Ijsselsteijn W et al (2000) Presence: concept, determinants and measurement. Proceedings of the SPIE 3959, pp 520–529

ISPR (2000) International society for presence research. Available via: http://www.temple.edu/ispr/frame_measure_t.htm. Accessed 30 Jan 2009

Jang D et al (2002) Analysis of physiological response to two virtual environments: driving and flying simulation. Cyberpsychol Behav 5(1):11–18

Klimmt C et al (2005) The value of reaction-time measures in presence research: empirical findings and future perspectives. Proceedings of the 8th international workshop on presence

Laarni J et al (2003) Using eye tracking and psychophysiological methods to study spatial presence, proceedings of the 6th international workshop on presence

Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 22(140):1–55

Lombard M et al (2000) Presence and television: the role of screen size, Hum Commun Res 26(1)

Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. Department of Broadcasting, Telecommunications and Mass Media, Temple University

Luchies C et al (2002) Effects of age, step direction, and reaction condition on the ability to step quickly. J Gerontol 57A(4):M246–M249

Nordahl R, Korsgaard D (2008) On the use of presence measurements to evaluate computer games. Proceedings of the 11th international workshop of presence, pp 174–177

Philip P et al (2004) Age, performance, and sleep deprivation. J Sleep Res 13(2):105–110

Reeves B et al (1992) Faces on the screen: pictures or natural experience? Paper presented to the mass communication division of the international communication association, Miami, FL

Reeves B et al (1993) New televisions: the effects of big pictures and big sound on viewer responses to the screen. Paper presented at the information systems division of the conference of the international communication association, Washington, DC

Rose L (2006) Hollywood's most expensive movies. Fobes.com. Available via: http://www.forbes.com/2006/12/18/movies-budget-expensive-tech-media-cx_lr_1214moviebudget.html. Accessed 29 Apr 2009

Rosmarin R (2006) Why gears of war costs $60. Fobes.com. Available via: http://www.forbes.com/technology/2006/12/19/ps3-xbox360-costs-tech-cx_rr_game06_1219expensivegames.html. Accessed 29 Apr 2009

Schlögl A et al (2002) Presence research and EEG. Proceedings of the 5th international workshop on presence

Slater M, Steed A (2000) A virtual presence counter. Presence Teleoper Virtual Environ 9(5):413–434

Steed A et al (2005) Breaks in presence as usability criteria. Proceedings of HCI international 2005

Sweetser P, Wyeth P (2005) Gameflow: a model for evaluating player enjoyment in games. Comput Entertain 3(3):3

Tan E (1995) Film-induced affect as a witness emotion. Poetics 23(1–2):7–32

Verrillo R (1962) Investigation of some parameters of the cutaneous threshold for vibration. J Acoust Soc Am 34(11):1768–1773

Verrillo R, Bolanowski S (1986) The effects of skin temperature on the psychophysical responses to vibration on glabrous and hairy skin. J Acoust Soc Am 80(2):528–532
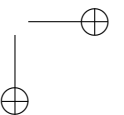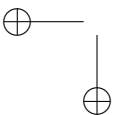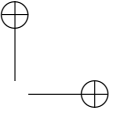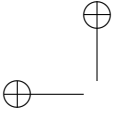
PAPER VII

R. Nordahl, and D. Korsgaard

On the use of adjustable distraction as a measure to
determine sustained attention during movie clips

On the use of adjustable distraction as a measure to determine sustained attention during movie clips

Rolf Nordahl
rn@media.aau.dk

Dannie Korsgaard
dkor05@media.aau.dk

Medialogy, Aalborg University Copenhagen, Lautrupvang 15, DK-2750 Ballerup

Abstract

Secondary task reaction time has been applied to measure the cognitive effects of movies, but has generally not produced convincing results when utilized to measure presence. We hypothesize that mixed results could be due to factors which prolong simple reaction time. This paper seeks to investigate the stability of a time independent alternative method (the adjustable distraction method) by testing it in relation to two other methods (heart rates and subjective intensity ratings) applied for measuring presence. In a comparison study 23 subjects were exposed to the three methods, with the aim of finding common and discrete patterns in the obtained results. The study showed that both the subjective ratings and the heart rates shared a common pattern, while the adjustable distraction method showed a discrete pattern. We speculate that this pattern emerged due to the great variance found in the participants' ability to perceive vibrations. In future research of the method we recommend extending it with a screening process to create uniform vibration perception thresholds.

Keywords--- Presence, immersion, attention, secondary task reaction time, adjustable distraction, vibration, measuring method, comparative test, movie clip, viewing session.

1. Introduction

The film media has a history reaching over a hundred years back [1] and yet we still struggle to understand its ability to engage, immerse and absorb us into alternative universes presented by the movies. A milestone towards understanding this intangible concept is to figure out how to quantify or measure it. One of the classic methods for evaluating features of movies has been the use of secondary task reaction time (STRT) [2,3,4]. The STRT method assumes that attention plays a part in this "media engagement". As such it attempts to measure how persistent a media product is in sustaining a user's attention. This is done by recording how long the user is in responding to a distraction (the secondary task), while experiencing the media. Through repetitive recordings it is possible to perform systematically evaluation of temporal changes in attention of a viewer during the playback of a movie. A couple of interesting findings have been done with this methodology. For instance the reaction time has been shown to increase if the amount of pans, zooms, cuts etc. is decreased [[3]]. This could indicate that excessive use of such narrative tools (e.g. pans and cuts) might have a repulsive effect on attention. It has also been shown that reaction time to random distractions increases in the second half of a viewing session [4], which might be because the viewer tends to forget the real world the longer they are exposed to an alternative experience offered by a media.

Among several researchers, the immersive effect, which media products can have on us, is known as *presence as immersion* [5]. Due to the STRT method's prominent role in prior media research it has been suggested as a possible mean to determine temporal changes in the level of presence during a media experience [6]. However experiments testing the STRT method in relation to presence have generally returned mixed results [7,8,9]. We hypothesize that these mixed results could be due to factors, which have an impact on simple reaction time (e.g. muscular tension [10], age [11,12,13], gender [13,14], fatigue [12] and breathing cycle [15]). To rule out these factors we have previously suggested an alternative method [16], which applies adjustable distraction as a measurement for sustained attention instead of reaction time. This paper will build upon the existing research on the use of the Adjustable Distraction method [16] to investigate whether it in relation to the STRT method can provide us with consistent results in relation to presence.

1.1. The adjustable distraction method

In a prior attempt to define a method for determining presence as immersion in media (e.g. games), the adjustable distraction (AD) method was suggested [[16]]. The AD method is still used to measure a media's persistency in sustaining the attention of a user. But instead of noting the reaction time, the strength of the distraction is repeatedly

increased or decreased in search of the minimum distraction strength at which the average media user shifts his or her attention from the media towards the distraction. The minimum strength of the distracting signal thereby becomes an indication of the media's persistency in sustaining users' attention.

The mechanism for adjusting the strength of the distraction has to be independent of time (e.g. the distraction could not gradually increase over time), because if time was used this would make all the time-related factors of the STRT method part of the AD method as well. Instead we came up with the idea to adjust the strength of the distraction across test participants and thereby obtain a general indication of "sustained attention" instead of an indication from each individual. To do so a particular mechanism was developed inspired by the binary search algorithm known from e.g. computer science [17]. To clarify how the mechanism works, for testing various distraction strengths, a fictional example will be used:

First the range of distraction strengths which should be evaluated must be determined before an experiment. For simplicity the range of signal strengths will be set from 0 to 10 in this example. 15 fictional participants have "volunteered" for this fictional experiment. The participants are divided into three groups with five in each. Now the first group watches a movie in turns and at the same point during the movie a signal of "strength 5" is send to each of the participants in the first group. Each of the five participants either does or does not react to the signal strength. If the majority (over 50%) of the first group reacted upon the signal strength the strength of the signal will decrease for the next group of participants. If the majority instead did not react upon the distraction strength the strength will increase for the subsequent group.

The exact strength presented to the next participants will be determined by the remaining span for strengths currently not tested. For instance we will assume that the majority of the first group reacted on "strength 5". This would suggest that there somewhere in the lower span, [strength 0; strength 5[, probably will be a strength at which the majority does not react. The best guess we can make at this point as to what strength this might be will be to take the strength-value in the middle of the lower span and test ("strength 2.5"). Therefore the next five participants will be exposed to this strength at the same point in the movie as the first five participants. Now we will assume that the majority of the second group did not react to the signal of "strength 2.5". Then it could be expected that there might be a strength in the upper span,]strength 2.5; strength 5[, at which the majority again would react upon and therefore the strength tested on the third group would be "strength 3.75". As the number of test rounds increments toward infinity the distraction strength will

become the borderline at which half of a population will not react, while the other half will react.

2. Methodology

To evaluate whether the results from AD method could be used as an indication of presence, we performed a comparative study. During the experiment a movie clip was shown to a range of participants, while three measuring methods were applied to collect information about the participants' feeling of presence. If all of these methods were measuring the same phenomenon a common pattern was expected in the collected data across the methods. If the results from any of the methods display a distinct pattern it will probably mean that the method is measuring another aspect of the experience. If this is so this study will contribute with considerations according to the authors observations on what the method might be measuring and how it might be altered to measure presence as immersion.

2.1. Test material

A seven minutes long movie clip from del Toro's "El Laberinto del Fauno" [18] was chosen as the test material for the experiment. Some of the arguments leading to this decision was the Spanish cast, which internationally was assessed as being relatively unknown. Thereby it was expected that less of the test participants would be affected by prejudices or certain views towards an actor acquired from seeing the actor performing in another movie. Another important feature of the selected clip was that its narrative structure was in accordance with the classical Hollywood model [1]. We felt this was an asset as the changing levels of suspense and action in this structure might result in a characteristic pattern of presence [19]. We hoped that this pattern would be easier to recognize across the results from all the measuring methods.

A quick description of the course of events in the movie clip is given in the following: A little girl has found a secret door leading to a room where a monster-like pale man is sleeping. While the girl is looking around the pale man wakes up and hunts her back to the door opening. The door shuts in front of her, but she manages to escape through another door in the ceiling, before the pale man reaches her. Now a cut is made to a scene where a group of people works through the woods.

During a pilot test the movie clip was shown to a couple of persons. According to their descriptions the intensity in the clip would almost constantly increase and peak at the scene where the girl attempts to escape. They also described how they suddenly perceived the pale man as appearing comical (instead of scary) at a point in the clip where it is revealed that the eyes of the pale man is located in the palms of his

hands. Due to these descriptions three events where chosen for evaluation:

Event 1: The sequence where the pale man suddenly appears comical.

Event 2: The part where the girl seeks to escape from the pale man.

Event 3: The beginning of the subsequent scene with the people walking in the woods.

It was expected that the level of presence would increase from event 1 to event 2, as the sudden comical appearance of the pale man might seem out of context with the eerie setting. Furthermore the feeling of presence was expected to decrease significantly from event 2 to event 3, because the subsequent scene will show a new location and new characters without proper introduction.

2.2. Data collection techniques

The three presence measuring methods compared were post-test paper and pencil questionnaires, Heart rate measurements and the AD method. Questionnaires were chosen as these are the common way of measuring presence [20]. Heart rates measurements were also interesting to include in the study as they measure temporal changes (like the AD method) and previously have shown convincing results when applied in tests of presence in scary [21] or violent media content [22]. Each method was to collect repeated measurements of the immersive presence felt during the three selected events in the movie clip.

The questionnaires consisted of three formulations created from the guidelines posted at the ISPR website [23]. It is here advised that presence is not directly addressed in subjective self-reports, but that descriptive synonyms should be applied instead [20]. As it was assessed to be too overwhelming and exhausting for the participants to answer a large range of similar questions containing different synonyms, it was decided to focus on finding a single descriptive term. As a result the three questions where formulated in the following way:

"How intense was the sequence where the pale man puts his hands (with his attached eyes) up in front of his face to look at the girl for the first time?"

"How intense was the sequence where the door closes in front of the girl, preventing her from escaping?"

"How intense was the sequence where the group of people is walking through the woods?"

Each of these questions where accompanied by a five optioned Likert-scale [24].

The AD method was utilized with vibration as the distracting stimulus. The vibration was created by a small pc speaker. The physical displacement of the speaker membrane (or the amplitude of the vibration) was then adjusted while the frequency of the speaker was kept constant at 250 Hz as the skin is most sensitivity to vibrations of this frequency [25].

2.5. Equipment

For the experiment a computer was set up and connected to the pc speaker (and an amplifier), through the sound card's output port. To transfer vibrations to the participants' skin a 5 x 50 mm stainless steel A4 bolt (non-magnetic material) was attached to the center of the pc speaker membrane. The speaker was placed onto a jack under a table in which a hole with a diameter of 12 mm had been drilled. The pointy bolt then had free passage through the hole and by raising the jack, the top of the bolt could then be placed with precision in relation to the upper side of the table top. In the other end of the table top a small USB numerical pad was placed and connected to the computer. The "Return"-key was marked to emphasize its significance above the rest of the buttons. The movie clip was displayed on a 32" LCD screen and the sound was provided through a set of muffled headphones connected to the screen. All the equipment was put up inside a sound proof room.

2.3. Procedure

The laboratory facilities at the campus of Aalborg University in Copenhagen were the setting of the experiment. 23 participants toke part in the experiment all between the ages of 20-30 years old. Prior to the test none of the participants had seen the particular movie clip before neither did they suffer from any heart disorders. It was also ensured that none of the participants knew Spanish beforehand (the spoken language in the movie). As it has been reported that low temperatures can have an effect on vibration perception [26] all participants where kept indoors for at least an hour before the experiment.

Each participant was asked to sit in a chair next to the table and then asked to put their left hand index finger on the "Return"-button on the numerical-pad. The participant was then asked to place his or her right hand on the surface of the table top just above the bolt-hole in a way that the bolt would hit the center of the first metacarpal of the thumb (the large pad in the palm just below the thumb). With their right hand in position the participant was instructed to keep it as still as possible. To ensure the same amount of pressure between the bolt and the skin of each participant, they were asked to hold

one end of an Ohm meter in the left hand while the other end was being attached to the bolt. When the jack was raised and made contact between the two ends of the circuit (the participants right hand and the bolt) the Ohm meter would display a reading. At this point the bolt would just barely touch the skin. To ensure full contact the bolt was then raised one millimeter.

The first phase of the experiment was used to find a participants perceivable limit of vibration (the vibration threshold) in a condition where the participant was unaffected by the test material. To do so the participant was asked to press the "Return"-button when he or she thought they could feel a vibration in their palm. Now the described mechanism for adjusting the strength of the stimulus in the adjustable distraction method was applied, but instead the reaction of the single participant was used to determine each test round. In this way the vibration signals send to the speaker (with random time intervals) was used to "close in" on each participant's threshold. During this phase the headphones emitted white noise to make sure that the vibrations could not be registered by their sound.

In the second phase the participant was equipped with electrodes in order to measure their heart rates. Instructions where again given to the participant regarding the pressing of the "Return"-button on the numerical-pad if they at any point thought they felt vibration in their palm. The playback of the movie clip was then started together with a computer program able to automatically (by the use of the adjustable distraction mechanism) send out signals to the speaker when a vibration should occur. After watching the movie clip each participant were asked to fill out the post-test questionnaire.

3. Results

It makes sense to list the results as graphs as the graphical relationships are helpful in detection of a pattern between the event-means. The means and standard deviations from the questionnaires are displayed in (Figure 1).
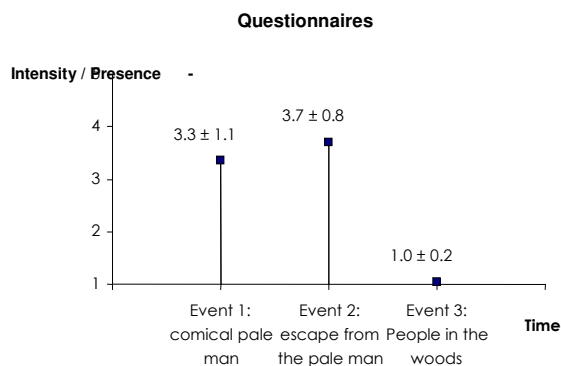
Questionnaires

Figure 1: Mean and standard deviation of the questionnaires intensity ratings at each of the three events.

The means and standard deviations of the heart rates recorded at each of the three events are likewise displayed in (Figure 2).
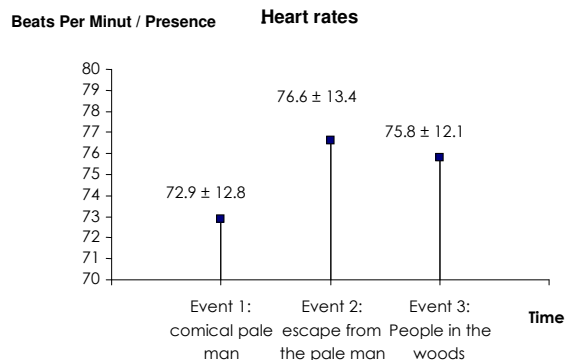

Heart rates

Figure 2: The mean and standard deviation calculated from the heart rate measurements taken at each of the three events.

The measuring of the participants' perceivable vibration threshold returned a mean value of 275.3 millivolts (mV) amplitude from peak-to-peak (p-p) of the sinusoid signal provided to the speaker. The standard deviation of the thresholds was rather high at 177.5 mV p-p. (Figure 3) illustrates the recordings from the AD method taken while the participants watched the movie clip.
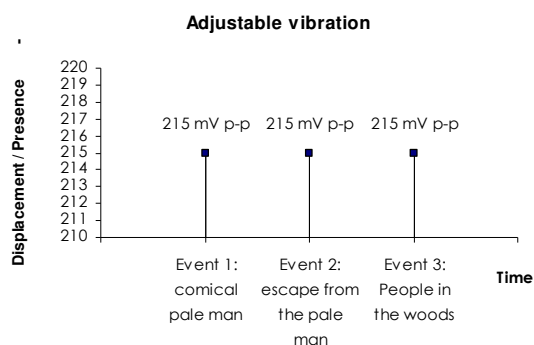

Adjustable vibration

Figure 3: Results returned from the AD search algorithm at the three events. Each result is the amplitude measured in milivolts (peak-to-peak) of the sinusoid signal send to the speaker.

4. Discussion

By comparing the three events ratings internally from the questionnaires (Figure 1) it is clear that the results are coherent with our expectations. The heart rate recordings (Figure 2) do show a similar relationship between event 1 and event 2 measurements, but event 3 is here resulting in a "medium" heart rate reading. We speculate that the latter is an indication of something else than presence. For instance it could be argued that the inconsistency between the scene with event 1-2 and the scene with event 3 could have left the participants wondering if the experiment was about to end. The participants might have asked themselves: "Did they forget to turn of the movie?" and similar questions in this period, resulting in a high heart rate due to this uncertain real-life situation. The results from the AD method (Figure 3) show a very different pattern. The search algorithm returned the same value (215 mV p-p) at each of the three events. This result was unexpected (as the results from a prior experiment with visual distraction returned promising results [16]), but the large variance in the participants' perceivable thresholds (standard deviation 177.5 mV p-p) might be a part of the explanation for this odd measurement. It is likely that a person with the ability to perceive a very weak vibration would experience a certain vibration strength stronger than another person which is only able to perceive high amplitude vibrations. If this observation holds true it would revoke the purpose of search algorithm, as the presented stimulus would be evaluated according to threshold rather than the level of presence experienced. To work around this issue in future tests a thorough screening process, to ensure uniform perceivable thresholds, might be included.
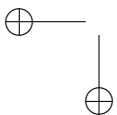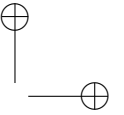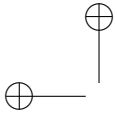
5. Conclusion

The experiment showed some similarities between the results from questionnaires and heart rate measurements, while the AD method returned results displaying a different pattern. From the great variance in the recorded vibration threshold collected from the participants', we believe that the AD method will continue returning doubtable results, if an included screening process is not performed. Such screening would lead to comparable thresholds between participants.

References

[1]   Bordwell, D., Thompsen, K. (2004). Film Art: An Introduction, 7th edn. McGraw-Hill Higher Education.

[2]   Geiger, S., Reeves, B. (1993). We interrupt this program... Attention for television sequences. *Human Communication Research*, 19, 368–387.

[3]   Thorson, E., Reeves, B., Schleuder, J. (1985). Message complexity and attention to television. *Communication Research*, 12, 427–454.

[4]   Lorch, E., Castle, V. (1997). Preschool Children's Attention to Television: Visual Attention and Probe Response Times. *Journal of Experimental Child Psychology*, 66(1), 111-127.

[5]   Lombard, M., Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Department of Broadcasting, Telecommunications and Mass Media*. Temple University.

[6]   Ijsselsteijn W et al. (2000). Presence: Concept, determinants and measurement. *Proceedings of the SPIE 3959*, 520–529.

[7]   Reeves, B. et al. (1992). Faces on the screen: Pictures or natural experience? Paper presented to the Mass Communication Division of the International Communication Association, Miami, FL.

[8]   Reeves, B. et al. (1993). New televisions: The effects of big pictures and big sound on viewer responses to the screen. *Paper presented at the Information Systems Division of the Conference of the International Communication Association*, Washington, DC.

[9]   Klimmt, C. et al. (2005). The Value of Reaction-Time Measures in Presence Research: Empirical Findings and Future Perspectives. *Proceedings of the 8th International Workshop on Presence*.

[10]   Etnyre, B., Kinugasa, T. (2002). Postcontraction influences on reaction time (motor control and learning). *Research Quaterly for Exercise and Sport*,73(3), 271-282.

[11]   Luchies, C. et al. (2002). Effects of age, step direction, and reaction condition on the ability to step quickly. *The Journals of Gerontology*, 57A(4), M246-M249.

[12]   Philip, P. et al. (2004). Age, performance, and sleep deprivation. *Journal of Sleep Research*, 13(2), 105-110.

[13]   Der, G., Deary, I. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey. *Psychology and Aging*, 21(1), 62-73.

[14]   Dane, S., Erzurumluoglu, A. (2003). Sex and handedness differences in eye-hand visual reaction times in handball players. *International Journal of Neuroscience*, 113(7), 923-929.

[15]   Buchsbaum, M., Callaway, E. (1965). Influence of respiratory cycle on simple RT. *Perceptual and Motor Skills*, 20, 961-966.

[16]   Removed due to review process.

[17]   Cormen, T. et. al. (2001). Introduction to Algorithms, 2nd edition, MIT Press.

[18]   del Toro, G. (2006). The Spanish movie "El laberinto del fauno". Official site available via: http://www.panslabyrinth.com

[19]   Freeman, J. et al. (2000). Using behavioural realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators and Virtual Environments*, 9(2).

[20]   Baren, J., Ijsselsteijn, W. (2004). Measuring Presence: A Guide to Current Measurement Approaches. Available via: http://www.informatik.umu.se/~jwworth/PresenceMeasurement.pdf

[21]   Jang, D. et al. (2002). Analysis of Physiological Response to Two Virtual Environments: Driving and Flying Simulation. *Cyberpsychology and behavior*, 5(1).

[22]   Cameron, P., Ballard, M. (2002). High Versus Low Aggressive Priming During Video-Game Training: Effects on Violent Action During Game Play, Hostility, Heart Rate, and

Blood Pressure. *Journal of Applied Social Psychology*, 32(12).

[23] ISPR (2000). International Society for Presence Research. Available via: http://www.temple.edu/ispr/frame_measure_t.htm

[24] Likert, R. (1932). A Technique for the Measurement of Attitudes, *Archives of Psychology*, 140.

[25] Verrillo, R. (1962). Investigation of Some Parameters of the Cutaneous Threshold for Vibration. *The Journal of the Acoustical Society of America*, 34(11).

[26] Verrillo, R., Bolanowski, S. (1986). The effects of skin temperature on the psychophysical responses to vibration on glabrous and hairy skin. *The Journal of the Acoustical Society of America*, 80(2).

[27] Reeves B et al. (1992). Faces on the screen: Pictures or natural experience? *Paper presented to the Mass Communication Division of the International Communication Association*, Miami, FL.

PAPER VIII

R. Nordahl, and D. Korsgaard

On the Use of Presence Measurements to Evaluate
Computer Games

PRESENCE 2008

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

Printed by
CLEUP Cooperativa Libraria Universitaria Padova
Padova 2008

Edited by Anna Spagnolli, Luciano Gamberini

ISBN: 978-88-6129-287-1

On the Use of Presence Measurements to Evaluate Computer Games

Rolf Nordahl, Dannie Korsgaard

Medialogy, Aalborg University Copenhagen, Ballerup, DK
{rn@media.aau.dk, dkor05@media.aau.dk}

Abstract

As the game industry expresses a growing demand for effective evaluation methods, it is worth investigating if the commonly used questionnaires can be replaced by alternative ways of measuring user experience in interactive environments. This paper describes an experiment where an existing presence measurement method was modified for use in computer game development. 39 subjects were part of the experiment, which was designed to test applicability of the adapted presence measuring method. Besides playing a game prototype, test participants were asked to press a button when a visual signal, triggered by an in-game event, would appear on the screen in the periphery of sight. Noting how strong the signal was is assumed to infer how strong the stimuli had to be in order to break the immersive presence. Results indicated that the adapted method with observations from the test is more useful, than questionnaires, in determining immersive presence at certain events.

Keywords--- **Computer games, Presence, Immersion, Measurement, Measurement Method, Game Development, User Experience**

1. Introduction

Game development budgets have over the years reached scales comparable with those of Hollywood film productions [1]. This entails that the development of a game is of high financial risk. In the corporate world, test and evaluation methods can be valuable tools to provide the foundation for informed predictions and decisions. Applying methods, which indicate how well a game translates into a positive user experience, might help the developers make the right design decisions, in order to create a commercially successful game.

For the aforementioned reasons, test and evaluation methods concerning the user experience in games have received an increasing amount of interest nowadays.

As a contribution to this discussion the authors wished to investigate whether measuring methods customized from the presence community could be adapted to evaluate games.

Our experiment attempts to adapt existing methods used to measure presence, in order to give game developers a tool to handle the intangibility of evaluating immersive capabilities of a game experience, according to the definition of presence as immersion in [3].

2. Measuring presence

In recent years, several techniques to measure presence have been proposed in the research community, and are well documented in the Presence research website. [20]

Questionnaires are one of the generally accepted and common ways to measure the effects of presence (p. 4)[2]. They are relatively easy to produce and work with, and their content has been elaborated over the years by several researchers.

One of the drawback of using questionnaires is the fact that subjects answer questions retrospectively. This can become a negative aspect as it relies on the memories of the participant, which may be incorrect and not correspond to his experience.

Even though questionnaires are easy to adapt, another issue rises when looking at the nature of computer games. Computer games are not usually linear constructs, in the sense that the content is interactive so the experience will very likely differ from one user to another. Therefore it can be difficult to match the questionnaires with each individual player's game experience. This is particularly problematic if the game developers wish to evaluate specific situations, or events, that occur in their game. The last and most significant shortcoming of questionnaires is that they rely on a subjective judgment from the user. Feelings, prejudices, personal impressions etc. can make an impact on this judgment, which may be undesirable.

2.1. Adapting an alternative

In [2], a collection of several lesser known presence measurement methods is collected. These methods are not directly designed to measure presence in games, but act as a good starting point for the production of such a method, that might also cope with the disadvantages of questionnaires. One of these alternatives is the selective attention measuring method (p. 40) [2], which exposes users to two tasks at the same time, and then attempts to measure presence in terms of which task is given the most attention by the user. In [4], this method was

---

[20] www.presence-research.org/

used but relied on the participants' memory to identify how attention was distributed between the two tasks, which once again is an encounter with the retrospective problem.

The interactive possibilities of computer games can be used to make one of the presented tasks into a measuring method that enables the participants to give feedback on-the-fly and thereby eliminating the occurrence of the described retrospective problem. By gradually exposing participants to a stronger stimulus, a recording of when each participant reacts on this (e.g. by pressing a physical button) could be regarded as a secondary task, which might be useful as documentation for the amount of experienced presence.

The shift of attention from the game to the stimuli can be described as a break in presence [5]. The intrusive stimulus is supposed to be activated at certain events during play. These events are to be chosen by the game developers and are likely to be some of the unique selling points that the developers wish to test.

3. The method

The stimuli used to provoke a break in presence could be visual-, sound- or even tactile signals. The important factor is that the signal can vary in strength, from undetectable to unavoidable. The stimuli signal which is chosen for this paper is a visual signal.

The assumption is that the player during a play session will treat the area around the center of the screen as the main point of visual attention, while the outermost areas of the screen will stay in the periphery of sight. As described in [3], the conceptualization of presence as immersion should entail that the test participant would shut out the world including perception of what is in the periphery of sight. If the intrusive stimulus is presented in the outermost areas one will be able to tell when a break in presence occurs.

4. Implementation

Placing a black border (a neutral zone) around the game visuals provided by the game makes up an edge, which separates two regions (Figure 1). The gradually changing visual simulation is made by altering the color from black to gray of a variable amount of pixels, formed as a circle, in a random position within the black border on the screen. The appearance of the circle or dot of a certain size should then either go unnoticed by or work intrusively and break the sense of presence and make the participant react by pressing a particular button on the mouse. As a participant can only either press the button or not, an extraordinary procedure is needed in order to make the method sensitive to different levels of presence.

Figure 5 The screen area divided in two: The actual game area and the neutral zone, in which the dot appears. Screen shot is taken from [6]

4.1. The binary search algorithm

When tracking improvements in game development, it is not enough only to know whether the player is immersed or not. To cope with this, a search pattern was implemented, which made it possible to test for different levels of presence. The algorithm works as following. A maximum diameter is set for the dot (typically the width of the black border) such that the dot size can vary from 0 to this maximum. Half of this value is set as the initial dot size for every event, which should be tested. Now an odd number (n) of test persons will play the game and when an event occurs a dot, of the initial size, will appear a random place in the neutral zone with the participant either reacting by pressing the key or not. The gathered binary results will now be evaluated, for each individual event, by the computer. If the majority (over 50%) reacted upon the dot of the initial size, the dot would increase, for the next test round, with half of the amount of pixels from the current size to the maximum dot size. Likewise, if the majority did not react upon the dot, the size would decrease with half of the pixels from the current size to zero. After the evaluation of the first test round, a new dot size for every event has been computed and a new round of n test persons can start playing the game with the appearance of the adapted dots. By always adjusting the dot sizes with either the upper or the lower half of the available range of untested sizes, eventually one size will be left when enough test rounds have been evaluated. This size is assumed be equal to the general amount of immersive presence experienced at a particular event. The accuracy of this measure will improve as n tends towards infinity.

5. Experiment

Before the experiment it was expected that the customized method would be better suited to measure immersive capabilities of specific events than questionnaires.

5.1 Procedure

To test the adapted measuring method, a prototype of a first-person shooter zombie-survival game was developed. 39 subjects participated to the experiment, 34 male and five female, between ages 20-30 and all full-time students. Participants were screened: prior to the test they were asked, if they enjoyed playing first-person shooters and whether they disliked violence in games. A negative answer to these questions would mean that they cannot be regarded as part of the game's target group and could therefore not be included in the experiment.

Test participants were divided into two groups to test the adapted method versus questionnaires. 26 participants, 25 male and 1 female were assigned to test the game using the secondary task measuring method (adapted group), while 13 participants, 9 male and 4 female were tested using questionnaires (regular group). Before playing the game, all participants were told about the objectives in the game and what the controls were. The adapted group was given additional instructions to press a mouse button when the dot would appear. Then the participants were asked to play the game at intervals of five persons per test round (n=5), after which they were asked to fill out the questionnaire.

The questionnaires contained seven questions, formulated according to [7] and meant to give an indication of the over-all immersive presence during the test session. In addition there were three questions regarding the immersion level at specific events. All of the questions were answered using a seven-optioned Likert scale [8]. As the game prototype evolved around the concept of survival horror, the gameplay was largely about stressing the player by exposing him/her to lots of attacking enemies. The main task was to survive for five minutes while fighting off an increasing number of zombies. A set of situations which could arise in connection with this task were, for example, events with many enemies near the player combined with little ammunition, which are assumed to create a more intense experience than events with just many enemies because of the limited means (ammunition) available to deal with the enemies. With that in mind, the questions concerning events were formulated as:

"How intense was the game when you had low/medium/high ammunition and were attacked by many/moderate amount of/few zombies?"

Since immersive presence [3] is a multi-dimensional concept, a problem was encountered at this point. Obviously it would be desirable to ask about as many of these psychological dimensions as possible, but doing so for each event would result in an impractically large questionnaire, so it was decided to focus on "intensity". This is problematic since the questionnaires will not measure the full effect of immersion, but rather just a small part of it.

It was expected that the level of immersion experienced would be ascending with the amount of zombies and descending with the ammunition available. For example, moments with low ammunition and many enemies nearby was assumed to be more intense, and thereby more immersive, than moments with few zombies and lots of ammunition. During the play sessions the three events were tested by the adapted group, where dots would appear when the criteria for the three events were met. Each dot could vary from 0 to 128 pixels in diameter and to avoid confusion, a dot would only appear the first time an event occurred for each player.

6. Results

After the test sessions, a mean value of immersive presence could be calculated from the seven over-all immersion-related questions in the questionnaires. This immersion mean and the standard deviation is displayed in the top of Table 1 for each group. The calculated means and standard deviations from the event-based questionnaire questions are listed in the rows below the immersion mean. All the means are based on a scale where 1 is low immersion/intensity and 7 is high immersion/intensity.

Parameters	Regular	Adapted
Immersion mean	4.49 ± 0.59	4.24 ± 0.97
Intensity, low ammo and many zombies	4.92 ± 1.62	5.22 ± 1.31
Intensity, medium ammo and moderate zombies	4.58 ± 1.38	3.70 ± 0.93
Intensity, high ammo and few zombies	3.77 ± 2.05	2.43 ± 0.95

Table 1 Displaying key values gathered from the questionnaires of the regular and adapted group

An unpaired t-test assuming equal variances between the immersion means from both groups returned $p < 0.4$. The results from the customized selective attention method are shown in Table 2 with respect to each event.

Event	Final dot size	Times evaluated	Remaining uncertainty
1. High ammo, few zombies	62 px	6	± 0 px
2. Medium ammo, moderate zombies	92 px	5	± 1 px
3. Low ammo, many zombies	24 px	4	± 3 px

Table 2 Results from the adapted measurement method

Assuming the adapted method is valid as a measurement of immersive presence, the size of the dot should correspond to the level of immersion experienced during the given event. "Times evaluated" differs between events because some events

occurred less often than others. As a limited amount of test participants were at disposal, the binary search algorithm did not manage to reach a single dot size value for all of the events. Therefore another parameter must be listed with the results. "Pixel uncertainty" refers to the maximum amount of pixels (px), which the dot size can vary at the arrived point in the search algorithm.

6.1 Discussion

By first comparing the two immersion means of each group, it is surprising that they come this close to each other due to the intrusive nature of the adapted method. Those who were exposed to the adapted measuring method experienced a slightly lower immersion level. But, a t-test ($p<0.4$) indicates that the difference is insignificant and that it could in fact be a result of the standard error.

Results from the event-questions in the questionnaire (Table 1), show that participants from both groups answered as expected. However it can be questioned if these answers are the outcome of the same logic thinking which created the expectation in the first place. This means that the questions might have been formulated in such a way, which induces the participants' to rely on logical thinking rather than what actually happened during the test. Results from the adapted method (Table 2) show a different pattern. Event three was expected to have the highest value, but is here assigned the lowest. At first glance these results might seem very irrational. However observations, during the test sessions, revealed that event three often got triggered when the player had no ammunition and was in a process of searching for new supplies. When the player is literally browsing the screen in order to find ammunition it could be expected that dots appearing at this point may be easier to detect. This suggests that the processes behind the results from the adapted method are more complex than first expected and that they might be giving a better indication than the questionnaires regarding the events.

As well as the adapted method is assumed to be able to measure presence at specific events, it might also be useful for getting an impression of the over-all amount of immersive presence. Calculating the average from the final dot sizes will results in a mean dot size. In this case it can be said that the [9]

general level of presence during the play sessions was at a point where it required a gray circle of 59 pixels in diameter to break it. This over-all value can serve as a mean for comparison between different stages of development or even between different computer game productions.
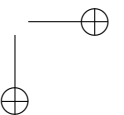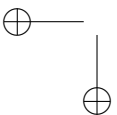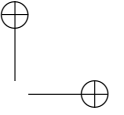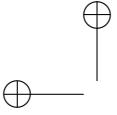
Conclusions

In this paper we proposed a technique to adapt presence measurements to computer games.

The questions regarding the intensity level of each event demonstrated how questionnaires can be hard to formulate in such a way to provide valid answers. Quantitative results from the adapted method accompanied by observations might be able to give some insights of the complex causes and effects contained within a prototype of a game and thereby assist in the improvement of the final computer game.

Based on these results it is suggested that further research and testing is conducted in determining the reliability of this and similar methods, which attempt to objectively measure the immersive capabilities of computer games.

References

[1]   R. Rosmarin. Why Gears of Wars cost $60. Forbes.com. URL: http://www.forbes.com/technology/2006/12/19/ps3-xbox360-costs-tech-cx_rr_game06_1219expensivegames.html

[2]   J. Baren, W. Ijsselsteijn. Measuring Presence: A Guide to Current Measurement Approaches. Deliverable 5, OmniPres project IST-2001-39237. 2004.

[3]   M. Lombard, T. Ditton. At the Heart of It All: The Concept of Presence. *Journal of Computer Mediated Communication*, 3. 1997. URL: http://jcmc.indiana.edu/vol3/issue2/lombard.html

[4]   R. P. Darken, D. Bernatovich, J. P. Lawson, B. Peterson. Quantitative Measures of Presence in Virtual Environments: The Roles of Attention and Spatial Comprehension. *Cyberpsychology and Behavior*, 2, 337-347. 1990.

[5]   M. Slater, A. Steed. A virtual presence counter. *Presence: Teleoperators and Virtual Environments,* 9, 413-434. 2000.

[6]   Warsow. URL: http://www.warsow.net

[7]   International Society for Presence Research (ISPR). URL: http://www.temple.edu/ispr/frame_measure_t.htm

[8]   R. Likert. A Technique for the Measurement of Attitudes. *Archives   of   Psychology*, 14, 1-55 1932.

LEGAL DOCUMENTATION

Co-author statements in connection with submission of PhD thesis

With reference to Ministerial Order no. 18 of 14 January 2008 regarding the PhD Degree ğ 12, article 4, statements from each author about the PhD student's part in the shared work must be included in case the thesis is based on already published articles.

The following documents are verifications and have been signed by the co-authors.

Co-author statement in connection with submission of PhD thesis

With reference to Ministerial Order no. 18 of 14 January 2008 regarding the PhD Degree § 12, article 4, statements from each author about the PhD student's part in the shared work must be included in case the thesis is based on already published articles.

Paper title: **Exploring sonic interaction design and presence: Natural Interactive Walking in Porto.**

Place of publication: Proceedings of Presence 2009
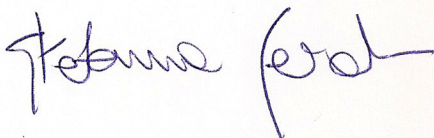
List of authors: Rolf Nordahl, Stefania Serafin, Federico Fontana

PhD student: Rolf Nordahl

Contribution: (%, text): 95
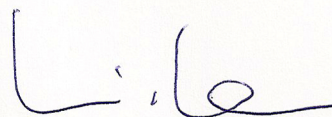
Signature, PhD student:

Signature, co-author

Stefania Serafin

Signature, co-author

Federico Fontana

Co-author statement in connection with submission of PhD thesis

With reference to Ministerial Order no. 18 of 14 January 2008 regarding the PhD Degree § 12, article 4, statements from each author about the PhD student's part in the shared work must be included in case the thesis is based on already published articles.

Paper title: **Sound synthesis and evaluation of interactive footsteps for virtual reality applications**
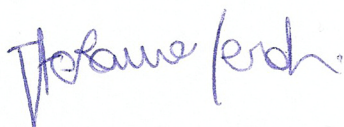
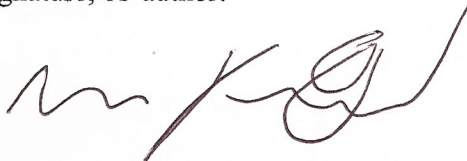Place of publication: Proceedings of IEEE VR 2010

List of authors: Rolf Nordahl, Stefania Serafin, Luca Turchet

PhD student: Rolf Nordahl

Contribution: (%, text): 85

Signature, PhD student:

Signature, co-author

Signature, co-author

Stefania Serafin

Luca Turchet

Co-author statement in connection with submission of PhD thesis

With reference to Ministerial Order no. 18 of 14 January 2008 regarding the PhD Degree § 12, article 4, statements from each author about the PhD student's part in the shared work must be included in case the thesis is based on already published articles.

Paper title: **On the use of adjustable distraction as a measure to determine sustained attention during movie clips**

Place of publication: Proceedings of Presence 2009, © International Society of Presence Research, 2009, ISBN 978-0-9792217-3-6

List of authors: Rolf Nordahl, Dannie Korsgaard

PhD student: Rolf Nordahl

Contribution: (%, text): 85

Signature, PhD student: Rolf Nordahl

Signature, co-author:

Dannie Korsgaard

Co-author statement in connection with submission of PhD thesis

Paper title: **On the Use of Presence Measurements to Evaluate Computer Games**

Place of publication: Proceedings of Presence 2008 © International Society of Presence Research, 2008, ISBN 978-88-6129-287-1
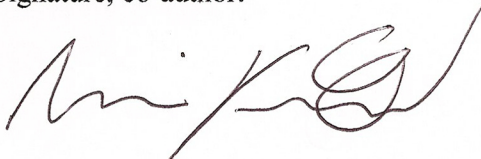
List of authors: Rolf Nordahl, Dannie Korsgaard

PhD student: Rolf Nordahl

Contribution: (%, text): 85

Signature, PhD student: Rolf Nordahl

Signature, co-author:

Dannie Korsgaard

Co-author statement in connection with submission of PhD thesis

With reference to Ministerial Order no. 18 of 14 January 2008 regarding the PhD Degree § 12, article 4, statements from each author about the PhD student's part in the shared work must be included in case the thesis is based on already published articles.

Paper title: **Distraction as a measure of presence: using visual and tactile adjustable distraction as a measure to determine immersive presence of content in mediated environments**

Place of publication: Virtual Reality © Springer-Verlag London Limited, 2009, DOI 10.1007/s10055-009-0140-3

List of authors: Rolf Nordahl, Dannie Korsgaard

PhD student: Rolf Nordahl

Contribution: (%, text): 85

Signature, PhD student: Rolf Nordahl

Signature, co-author:

Dannie Korsgaard