Sound Synthesis and Evaluation of Interactive Footsteps and Environmental Sounds Rendering for Virtual Reality Applications

Nordahl, Rolf; Turchet, Luca; Serafin, Stefania

Published in:
I E E E Transactions on Visualization and Computer Graphics

DOI (link to publication from Publisher):
10.1109/TVCG.2011.30

Publication date:
2011

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

Sound Synthesis and Evaluation of Interactive Footsteps and Environmental Sounds Rendering for Virtual Reality Applications

Rolf Nordahl, Luca Turchet, *Student Member*, *IEEE*, and Stefania Serafin

Abstract—We propose a system that affords real-time sound synthesis of footsteps on different materials. The system is based on microphones, which detect real footstep sounds from subjects, from which the ground reaction force (GRF) is estimated. Such GRF is used to control a sound synthesis engine based on physical models. Two experiments were conducted. In the first experiment, the ability of subjects to recognize the surface they were exposed to was assessed. In the second experiment, the sound synthesis engine was enhanced with environmental sounds. Results show that, in some conditions, adding a soundscape significantly improves the recognition of the simulated environment.

Index Terms—Sound and music computing, walking, surface simulation, soundscape rendering.

1 Introduction

In the field of virtual reality, the development of efficient yet accurate simulation algorithms, together with improvements in hardware technology, has boosted research on auditory display and physically based sound models for virtual environments (VEs) [39], [32], [9].

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 [36], [33], sound quantity and quality of auditory versus visual information [6] and 3D sound [13], [40]. Recent studies have investigated the role of auditory cues in enhancing self-motion and presence in virtual environments [20], [18], [38].

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.

These 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. Typically, these 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 [7]. Among these algorithms, the sounds of people walking on different surfaces were simulated [8]. A similar algorithm was also proposed in [12], where physically informed models reproduced several aggregate surfaces. Procedural sound synthesis of walking has also been recently described in [11].

When exploring a place by walking, one can identify two main categories of sound: one's own footsteps and the surrounding soundscape. In the movie industry, footstep sounds represent important elements. 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 [5]. Studies on soundscapes originated from the work of R. Murray Schafer [34]. Among other ideas, Schafer proposed soundwalks as empirical methods for identifying soundscapes for specific locations. During a soundwalk people are asked to move in a specific location, paying attention to 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 [4]. By synthesizing technical, aesthetical, and humanistic considerations, the authors describe the field of aural architecture and its importance in everyday life.

The study of human perception of locomotion sounds has addressed several properties of walking sound sources: the gender [22], [15], and posture of a walker [29], the emotions of a walker [15], the hardness and size of the shoe sole [15], and the ground material [16].

Such studies have been concerned only with recognition of sounds in an offline scenario, where subjects were asked to listen to some sounds and classify them.

Concerning interactive footwear, previous work such as research performed by Paradiso and coworkers [28], [3] consisted of designing shoes augmented with sensors used to control footstep sounds. A smaller number of examples,

• *The authors are with the Department of Architecture, Design and Media Technology, Aalborg University Copenhagen, Lautrupvang 15, 2750 Ballerup, Denmark. E-mail: {rn, tur, sts}@media.aau.dk.*

Manuscript received 15 Aug. 2010; revised 21 Nov. 2010; accepted 11 Dec. 2010; published online 31 Jan. 2011.
Recommended for acceptance by K. Kiyokawa, G. Klinker, and B. Lok.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCGSI-2010-08-0184.
Digital Object Identifier no. 10.1109/TVCG.2011.30.

such as recent work of Nordahl [23] and Law et al. [21], have even aimed to provide multimodal cues linked to footsteps events in such environments.

In [35], we presented the first prototype of a system that is able to synthesize in real-time the sound of footsteps on different materials. This interactive system was evaluated in a between-subjects experiment, where it was compared to a recognition task including recorded and synthesized offline sounds. Results showed that subjects were able to recognize most of the synthesized surfaces with high accuracy. Similar accuracy was also noticed in the recognition of real recorded footstep sounds, which was an indication of the success of the proposed algorithms and their control [25].

In this paper, we are interested in understanding whether the addition of a soundscape enhances the recognition of the simulated surfaces. Our hypothesis is that context plays an important role in the recognition of the material a person is stepping upon.

The faithful simulation of auditory cues is often a neglected aspect in virtual reality research. In this paper, we both introduce a physically based sound synthesis engine which simulates footstep sounds, and we describe three experiments which were performed in order to achieve a better understanding of the role of virtual footstep sounds and the surrounding soundscape in simulating an auditory space.

We begin by describing in Section 2, the sound synthesis engine developed and in Section 3 its relative control. In Section 4, we then describe the first experiment which assessed the ability of subjects to recognize the synthesized surfaces, and we then present in Section 5 a second experiment where the simulation of footsteps is enhanced by the addition of environmental soundscapes.

2 The Sound Synthesis Engine

We developed a physically based sound synthesis engine that is 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 a recurring idea in auditory display research during recent decades [14]. 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, snow, or sand.

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 [19], [39]. These sounds are typically short in duration, with a sharp temporal onset and relatively rapid decay. A good physically motivated synthesis for solid objects is modal synthesis [1], [39], where a vibrating object is modeled by a bank of harmonic oscillators excited by and external stimulus. 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 response is modeled as a bank of filters with impulse response $h(t) = \sum_i a_i e^{-b_i t} \sin(2\pi f_i t)$, 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. Frequencies and decay rates of the modes are determined by the geometry and material properties of the object, and the gains of the modes depend on contact location on the object. A footstep sound may be considered to cause multiple microimpacts between the sole of a shoe, i.e., an *exciter*, and a floor, i.e., a *resonator*. This interaction can be either discrete, as in the case of walking on a solid surface, or continuous, as in the case of a foot sliding across the floor.

In the simulation of discrete impacts, the external excitation is brief and has an unbiased frequency response. The interaction is modeled by a Hunt-Crossley-type interaction where the force, f, between two bodies, combines hardening elasticity and a dissipation term [17]. Let x represent contact interpenetration and $\alpha > 1$ be a coefficient used to shape the nonlinear hardening, the special model form we used is

$$f(x, \dot{x}) = -kx^{\alpha} - \lambda x^{\alpha}\dot{x} \quad \text{if } x > 0, \quad 0 \text{ otherwise.}$$

The model described was discretized as proposed in [2].

If the interaction called for slip, we adopted a model where the relationship between relative velocity v of the bodies in contact and friction force f is governed by a differential equation rather than a static map [10]. Considering that friction results from a large number of microscopic damped elastic bonds with an average deflection z, a viscous term $\sigma_2 v$, and a noise term $\sigma_3 w$, to represent roughness, we have

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

The force specified by these models is applied to a virtual mass which produces a displacement signal that is then processed by a linear shaping filter intended to represent the resonator.

2.2 Aggregate Surfaces

The synthesis algorithms just described are directly applicable only to cases in which the ground surfaces consist of a solid body. To simulate footsteps onto aggregate ground materials such as sand, snow, gravel, a common temporal process based on a stochastic particle model is needed. Different particles do not have to be modeled explicitly, but only the probability that particles will create some noise is simulated.

The simulation is therefore characterized by energies and transition times that depend on the characteristics of the system and the amount of power it absorbs while changing configuration. The simulation dynamically captures macroscopic information about the resulting composite system through time [12].

Specifically, the simulation is performed in terms of the probabilistic distribution of the energies E of the short transients, which can be assumed to follow a power law $p(E) \propto E^{\gamma}$, and a model of the temporal density $N(t)$ of transients as a stationary Poisson process, under the assumption that the intertransient event times τ are

Fig. 1. A screenshot of the graphical user interface for the developed sound synthesis engine.

assumed to be independent: $P(\tau) = \lambda e^{\lambda\tau}$ [27]. The value of γ determines the type of noise produced by the process. 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.

Several models of this general type have been developed in order to mimic the sound of a footstep onto aggregate grounds [8], [12], [26].

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 further enhanced by using some reverberation. Reverberation was implemented by convolving in real-time the footstep sounds with the impulse response recorded in different indoor environments.

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] A screenshot of the final graphical user interface can be seen in Fig. 1.

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. In the synthesis of aggregate materials, parameters such as intensity, arrival times, and impact 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. These controls enable the sound designer to choose foot-ground contact sounds from a particularly rich physically informed palette.

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 aggregate models having the same density, duration, frequency, and number of colliding objects.

A pilot test was run to ascertain that 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.

The amplitude of the different components were also weighted, according to the same contribution present in the corresponding real sounds. Finally, a scaling factor was applied to the volumes of the different components. This was done in order to recreate a sound level similar to the one happening during a real footstep on each particular material.

3 Controlling the Sound Synthesis Engine

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 developed sound synthesis engine is controlled in real time by extracting the GRF while subjects perform the act of walking. Specifically, as can be seen in Fig. 2, subjects are asked to walk on a medium density fiberboard (MDF) which is 2.5×2 m in size and 1 cm thick. Four microphones are placed on the sides of the fiberboard in a square configuration. Specifically, we used four Shure BETA 91,

1. www.cycling74.com.
2. www.puredata.org.
3. http://puredata.info/Members/thomas/flext.

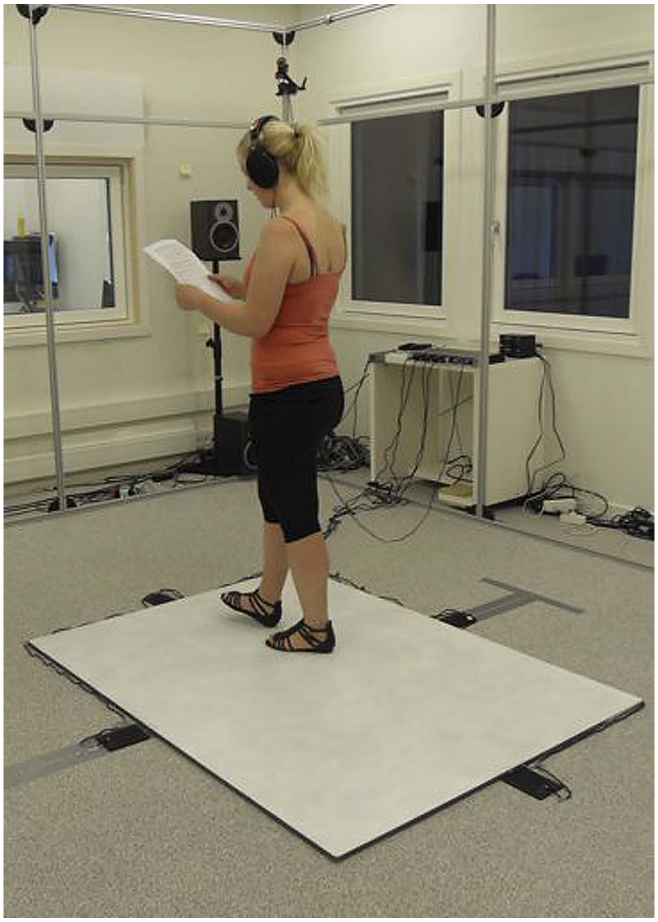

Fig. 2. A subject using the interactive footsteps synthesizer. Notice the MDF board surrounded by the four microphones.

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 footstep sounds. A multichannel Fireface 800 soundcard receives audio signals from the microphones, and transmits it via Firewire to a Macbook Pro laptop. Here, their GRF is estimated from the input signal. In order to perform a more precise estimation, among the four footstep sounds detected by the four microphones, we choose the one which is higher in amplitude.

To calculate the GRF, we use a simple nonlinear low-pass filter proposed by Cook in [30]:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1),$$

where

$$b = \begin{cases} b_{up} & \text{if } |x(n)| > e(n-1) \\ b_{down} & \text{otherwise.} \end{cases}$$

An example of a footstep sound and its corresponding GRF is shown in Fig. 3. The extracted GRF is used to control the force parameter of the different sound synthesis engines. For example, as shown in Fig. 4, the extracted GRF is used to control in real time the force parameter of the impact model. Finally, the synthesized sounds, as well as the soundscapes, are provided to the user by means of a set of Beyerdynamic DT-770 headphones. Fig. 5 shows a schematic representation of the different hardware components used.

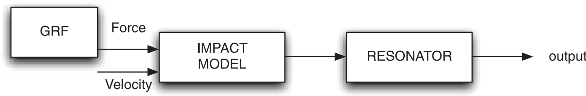

Fig. 4. Block diagram representing the algorithm and control of solid material sounds.

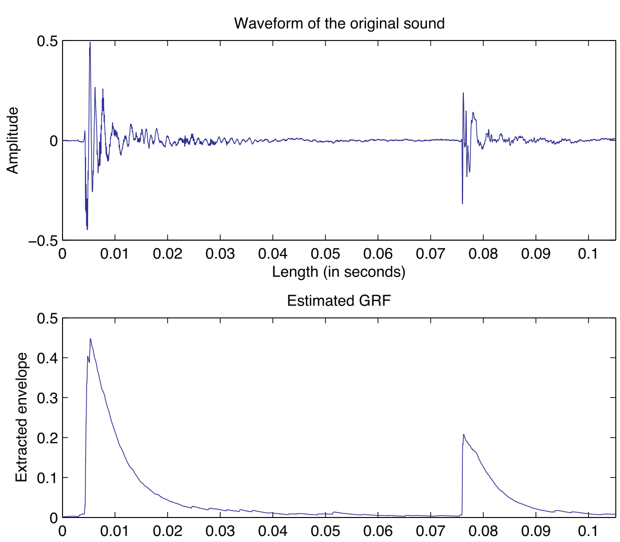

Fig. 3. A footstep sound (top) and the corresponding calculated GRF (bottom).

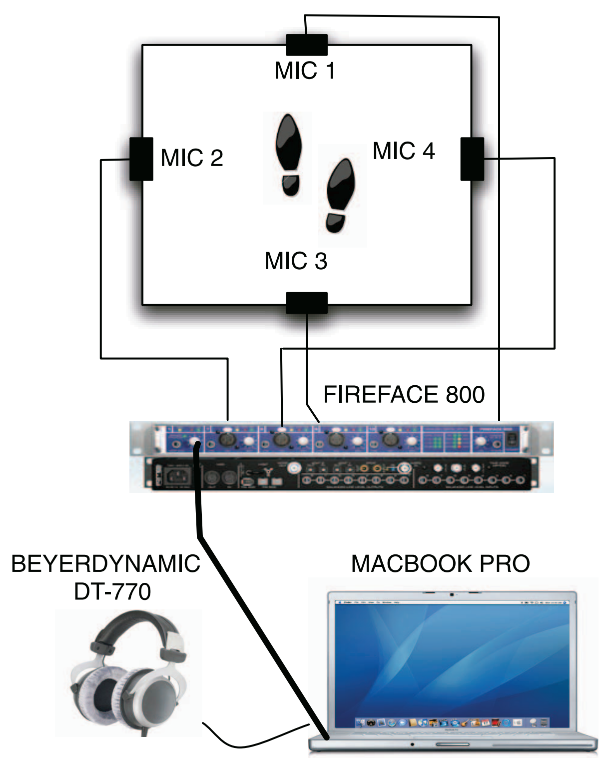

Fig. 5. Hardware components of the system developed: microphones, multichannel soundcard, laptop, and headphones.

TABLE 1
Confusion Matrix for Experiment 1: Recognition of Synthesized Footstep Sounds

	BS	GL	DR	SW	HG	UB	DL	WD	CW	MT	WR	CR	MR	FS	CC	PD	WT	CP	—
BS	15	2		5			1							2				1	4
GL		21	2			1	1							4					1
DR		1	3	2		6	6		1					10					1
SW				24	1									4					1
HG	2	7	3	1	0	3	7			2									5
UB		1	3	1		19	1							3				1	1
DL	1	3	5			5	12							4					
WD			1	2				14		1						1	1		10
CW								1	28					1					
MT								1		24				1	2				2
WR									3	11	6				7				3
CR												28			1				1
MR								1					25	1	1				2

Legend:

WD	wood	CW	creaking wood	SW	snow	UB	underbrush
—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
MT	metal	HG	High grass	DL	dry leaves	CC	concrete
DR	dirt	PD	puddles	WT	Water	CP	carpet
WR	wood reverb	MR	metal reverb	CR	creaking+ reverb		

Informal tests with several participants during the development stage show that no perceived latency is experienced. Moreover, the amplitude of the synthetic footsteps is high enough to mask the one of the real footsteps, in such a way that subjects perceive only the synthetic ones. The kind of shoes which the subjects are wearing affects the overall shape of the GRF, which is mapped to the amplitude of the sound synthesis engine. Therefore, the kind of shoes worn by the subjects is reflected in the synthesis engine. This is is a desired feature of the engine, since it matches the real world experience where different shoes produce different sounds when people are walking on the same surface.

4 Experiment 1

In the first experiment, we investigated the ability of subjects to recognize the different walking sounds they were exposed to. In this experiment, subjects were asked to recognize the sounds in an active setting, i.e., by using the developed interactive system.

4.1 Method

The sounds provided during the experiment were synthesized sounds generated in real time while subjects were walking using the interactive system described in the previous section.

Participants were exposed to 26 trials, where 13 stimuli were presented twice in randomized order. The stimuli consisted of footstep 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, footstep sounds on wood, creaking wood, and metal were enhanced adding a certain amount of reverberation.

4.1.1 Participants

Fifteen participants (six men and nine women), aged between 19 and 29 ($\text{mean} = 22.13, \text{standard deviation} = 2.47$), took part to the experiment. All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

The shoes used by subjects were sneakers, trainers, boots, and other kinds of shoes with rubber sole. The participants took on average 24 minutes to complete the experiment.

4.1.2 Setup

All experiments were carried out in an acoustically isolated laboratory where the setups were installed. The setup consisted of the interactive system, and the participants were asked to use it in order to generate the footstep sounds in real time (see Fig. 2).

4.1.3 Task

Participants were asked to wear a pair of headphones and to walk in the area delimited by the microphones. They were given a list of different surfaces to be held in one hand, presented as nonforced alternate choice. The list of surfaces presented to the subjects is outlined in the first row of Table 1. It represents an extended list of the surfaces the subjects were exposed to.

At the end of the experiment, subjects were asked some questions concerning the naturalness of the interaction with the system.

4.2 Results

Table 1 shows the confusion matrix which displays the results of Experiment 1. The matrix shows information concerning actual classifications performed by the subjects. The first row represents the list of materials the subjects could choose from, while the first column represents the actual stimuli subjects were exposed to. The choice of having a wider list of materials was made to avoid as much as possible that subjects simply guessed a choice.

From this table, it is possible to notice how surfaces such as snow, creaking wood with and without reverberation, gravel and metal with reverberation were correctly recognized in a high number of instances. Recognition of surfaces such as dirt plus pebbles, high grass, and wood appeared to be wrong most of the times. 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 versus concrete, snow versus frozen snow, dry leaves versus forest underbrush) rather than to different categories (e.g., wood versus water, wood versus gravel, and metal versus 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 (no tangible differences were found for the creaking wood). Overall, recognition rates are similar to those measured on recorded footstep sounds [25].

5 Experiment 2

The first experiment does not take into account the fact that when walking in a space, either indoor or outdoor, subjects are exposed not only to the sounds of their own footsteps, but also to sounds of the environment. Our hypothesis is that environmental sounds play an important role in the recognition of the surface subjects are stepping upon. We tested our hypothesis by conducting an experiment whose goal was to investigate the ability of subjects to recognize the different walking sounds they were exposed to under three conditions: without an accompanying soundscape, with a coherent soundscape, and with an incoherent soundscape. In this experiment, we define coherent soundscape as a soundscape that matches the expectations of the subjects. Specifically, soundscapes of the following environments were designed:

1. A beach and seaside during the summer;
2. A courtyard of a farm in the countryside;
3. A ski slope;
4. A forest; and
5. A park during the fall.

The specific content of the soundscapes was designed according to the indications given by subjects answering to a questionnaire. Precisely, ten subjects, chosen among those not participating to the final experiment, were asked to imagine which sounds could occur in the above-mentioned environments. Specifically, subjects were asked the following question: "Imagine that you are right now in a forest: which sounds do you think you would hear?" In this particular environment, subjects indicated sounds like trees, birds, and different animals. Among the answers provided, we chose those which were stated by more than one subject, and collected a corresponding sound material using recordings of real sounds.

The most occurring named sound events were used as sonic elements to build the final soundscape. The sounds were chosen among those available both on the Hollywood Edge sound effects library[4] and on the Freesound website.[5] The chosen sounds were using the sound editor Adobe Audition 3.[6] Soundscapes were designed with the goal of providing a clear idea of the designed environment already from the first seconds. As an example, in the soundscape of the forest animals characteristic of that environment, such as the cuckoo, could be heard, and the density of natural sounds is higher as compared to the soundscape of a park. The footstep sounds provided during the three conditions were synthesized sounds generated in real time while subjects were walking using the interactive system described previously. The soundscapes were audio files played in background independently from the subjects movements. The amplitudes of both footsteps and soundscapes were set by empirical investigation.

5.1 Method

A between-subject experiment with the following three conditions was conducted:

1. Condition 1: footstep sounds without soundscapes.
2. Condition 2: footstep sounds with coherent soundscapes.
3. Condition 3: footstep sounds with incoherent soundscapes.

The first condition was meant as a validation of the results of experiment 1, but using a smaller set of surfaces.

Subjects were exposed to 10 trials in conditions 1 and 2, where five stimuli were presented twice in randomized order. The stimuli in condition 1 consisted of footstep sounds on the following surfaces: beach sand, gravel, snow (in particular deep snow), forest underbrush (a forest floor composed dirt, leaves and branches breaking), dry leaves. In condition 2, the stimuli consisted of the same footstep sounds provided in condition 1 with in addition the corresponding coherent soundscape described in Section 5.

In condition 3, subjects were exposed to six stimuli presented twice in randomized order. The stimuli consisted of footstep sounds on the surfaces beach sand, snow, and forest underbrush with, in addition, an incoherent soundscape. As an example of incoherent soundscape, in presence of footstep sounds on beach sand, the provided soundscape corresponded to that of a landscape with snow (i.e., a ski slope) and forest underbrush (i.e., the forest environment). The choice of using sand, snow, and forest underbrush as surfaces was justified by the fact that we wanted to select three very distinct outdoor landscapes (beach, mountain, and forest).

5.1.1 Participants

Forty-three participants were divided into three groups to perform the three conditions in a between-subjects experiment ($n = 15$, $n = 15$, and $n = 13$, respectively). The three groups were composed, respectively, of 11 men and 4 women, aged between 21 and 28 ($\text{mean} = 23.67$, $\text{standard deviation} = 2.12$), 8 men and 7 women, aged between 19 and 38 ($\text{mean} = 24.67$, $\text{standard deviation} = 5.97$), and 6 men and 7 women, aged between 21 and 30 ($\text{mean} = 24$, $\text{standard deviation} = 3.1$). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

During the experiment, the shoes used by subjects were sneakers, trainers, boots, and other kinds of shoes with rubber sole. The participants took on average 11, 13, and 16 minutes in average for conditions 1, 2, and 3, respectively.

5.1.2 Setup

The experiment was carried out in the same acoustically isolated laboratory as of the previous experiment, where the MDF delimited by four microphones placed in a square configuration was installed (see Fig. 2).

4. http://www.hollywoodedge.com
5. http://www.freesound.org/
6. http://www.adobe.com/products/audition/

TABLE 2
Results of Condition 1, Experiment 2: Recognition of the Surfaces without Soundscapes

	% Correct answers	% Wrong answers	% "I don't know"	Realism	Quality
Beach Sand	50.00	46.67	3.33	5.20	4.72
Gravel	83.33	6.67	10.00	5.20	4.63
Snow	73.33	26.67	0.00	5.30	5.12
Forest underbrush	40.00	50.00	10.00	3.50	4.19
Dry Leaves	16.67	63.33	20.00	4.40	3.92

TABLE 3
Results of Condition 2, Experiment 2: Recognition of the Surfaces with Coherent Soundscapes

	% Correct answers	%Wrong answers	% "I don't know"	Realism	Quality	% Correct soundscape
Beach Sand	86.67	10.00	3.33	5.15	5.52	93.33
Gravel	86.67	13.33	0.00	5.31	5.40	86.67
Snow	80.00	20.00	0.00	6.17	6.07	83.33
Forest underbrush	73.33	26.67	0.00	4.91	5.03	100.00
Dry Leaves	30.00	70.00	0.00	4.44	4.50	96.67

TABLE 4
Results of Condition 3, Experiment 2: Recognition of the Surfaces with Incoherent Soundscapes

Material	Soundscape	% Correct answers	% Wrong answers	% No idea	Realism	Quality	% Correct soundscape
Beach Sand	Forest	38.46	57.70	3.84	4.00	5.16	88.46
Beach Sand	Ski slope	15.38	69.24	15.38	3.75	4.32	38.46
Snow	Forest	50.00	42.31	7.69	5.29	5.35	96.15
Snow	Beach	50.00	46.16	3.84	4.77	5.00	88.46
Forest underbrush	Beach	38.46	53.85	7.69	4.80	4.96	65.38
Forest underbrush	Ski slope	30.76	65.40	3.84	4.30	4.61	46.15

5.1.3 Task

During the experiment, subjects were asked to wear a pair of headphones and to walk on the MDF in the area delimited by the microphones. They were given a list of different surfaces to be held in one hand, presented as nonforced alternate choice. The list included a range of materials wider than those presented in the experiment. As before, the choice of having a wider list of materials was made to avoid as much as possible that subjects simply guessed a choice.

During the act of walking, subjects listened simultaneously to footstep sounds on a different surface according to the stimuli presented. The task, common to the three conditions, consisted of orally answering 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:
   a. beach sand,
   b. gravel,
   c. dirt,
   d. snow,
   e. high grass,
   f. forest underbrush,
   g. dry leaves,
   h. wood,
   i. creaking wood,
   j. metal,
   k. carpet,
   l. concrete,
   m. frozen snow,
   n. puddles,
   o. water, and
   p. 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 ($1 = \text{low realism}$, $7 = \text{high realism}$).
3. Evaluate the quality of the sound on a scale from 1 to 7 ($1 = \text{low quality}$, $7 = \text{high quality}$).

In conditions 2 and 3, participants were also asked to recognize what was the environment they were walking on.

They were informed that they could choose the same material more than once and that they were not forced to choose all the materials in the list. In addition, they could use the interactive system as much as they wanted before providing an answer. After moving to the next stimulus, subjects could not change their answers to the previous stimuli.

At the end of the experiment, the subjects were also given the opportunity to leave an open comment on their experience interacting with the system.

5.2 Results

The collected answers were analyzed and compared between the three conditions. Results are shown in Tables 2, 3, and 4.

The first noticeable element emerging from the three tables is that the use of the interactive system in the condition of coherent soundscapes gave rise to a higher recognition rate and a higher evaluation of realism and quality of the proposed sounds, compared to the conditions with no soundscapes and with incoherent soundscapes. Concerning

TABLE 5
Confusion Matrix for Condition 1, Experiment 2: Recognition of Footstep Sounds without a Soundscape

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	15	2	6	2	2	2										1
GL		25			1								1			3
SW			22	1	1			6								
UB			4	12			1	10								3
DL		9		4	5			6								6

Legend:	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

TABLE 6
Confusion Matrix for Condition 2, Experiment 2: Recognition of Footstep Sounds with a Coherent Soundscape

	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	26	1	1	1												1
GL		26			3	1										
SW	2		24					4								
UB		1	2	22	1			4								
DL	6			9	9			6								

Legend:	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

TABLE 7
Confusion Matrix for Condition 3, Experiment 2: Recognition of Footstep Sounds with an Incoherent Soundscape

	Soundscape	BS	GL	SW	UB	DL	HG	DR	FS	WD	CW	MT	CC	PD	WT	CP	—
BS	Forest	10		2	5	3	1	1	3								1
BS	Ski slope	4	1	5		1	5	3	1					1	1		4
SW	Forest		5	13	1	1			4								2
SW	Beach	2	6	13	2				2								1
UB	Beach		3	4	10	2			5								2
UB	Ski slope		1	1	8	3	1	2	8							1	1

Legend:	WD	wood	CW	creaking wood	SW	snow	UB	underbrush
	—	don't know	FS	Frozen snow	BS	beach sand	GL	Gravel
	MT	metal	HG	High grass	DL	dry leaves	CC	concrete
	DR	dirt	PD	puddles	WT	Water	CP	carpet

the percentages of correct answers, they are higher for condition 2 compared to condition 1, and this is true for every surface. The analysis by means of a chi-square test reveals that differences are statistically significant for beach sand ($p < 0.01$) and forest underbrush ($p = 0.019$). Especially in the condition where subjects were exposed to a stimulus representing a footstep on a forest underbrush, the soundscape was extremely important in facilitating the recognition. As a matter of fact, the soundscape, contained several elements characteristics of a forest, such as birds singing, which represented important cues for the subjects. It is particularly interesting to notice that overall adding a soundscape enhances the recognition factor, and this is especially noticeable for those situations where the recognition was rather low without a soundscape.

Similarly, the percentages of correct answers are higher for condition 2 compared to condition 3 for each surface. Differences are statistically significant for beach sand ($p < 0.01$), snow ($p = 0.0144$), and forest underbrush ($p < 0.01$). Furthermore, the percentages of correct answers are higher for condition 1 compared to condition 3, for each surface, but the differences are not statistically significant.

The analysis of the wrong answers reveals that in all the experiments none of the presented surfaces was recognized as a solid surface. This means that all subjects were able to identify at least the nature of the surface, which was an expected feature of the simulations. After the experiment, several subjects observed that the simulated footstep sounds were perceived as very similar, and therefore hard to recognize and distinguish from the list provided.

It is interesting to examine what happens when subjects are exposed to soundscapes which are incoherent, as shown in Table 4. In this situation, we consider as correct the answer provided when subjects recognize the surface they are walking on, and not the soundscape. As it can be noticed, the percentage of correct answers is quite low. As expected, adding an incoherent soundscape creates a stronger context which often confuses the subjects. This can be observed, for example, in the case of beach sand footsteps which were rendered together with a forest soundscape and a ski slope soundscape. The recognition rate is higher in the first case than in the second, where several subjects confused sand with snow. The answers for the three conditions are outlined in the confusion matrices shown in Tables 5, 6, and 7, respectively. As before, the first row represents the list of materials the subjects could choose from, while the first column represent the actual stimuli subjects were exposed to.

From the matrices, it can be noticed how the subjects' recognition varies from condition 1 to condition 2. As an example, the second row of the matrix illustrates the number of subjects which recognized the beach sand surface, with Table 5 and without Table 6 a soundscape. The role of the soundscape to enhance the recognition is clearly noticeable.

Table 7 illustrates the confusion matrix for condition 3, i.e., when incoherent soundscapes are presented to the subjects. In this situation, it is clearly noticeable how the nature of the soundscape plays an important role. Moreover, it can be noticed how the incoherent soundscape is in most situation predominant, in the sense that subjects tend to judge the surface they are stepping upon more listening to the soundscape than listening to the actual surface. On the other hand, even if subjects are not able to recognize the surface they are stepping upon, they never confuse its nature, in the sense that they never select a solid surface when exposed to an aggregate one.

In addition, Tables 2, 3, and 4 show the degree to which participants judged the realism and quality of the experience. The degree of realism was calculated by looking only at that data from correct answers, i.e., when the surfaces were correctly recognized. This choice was performed since we were interested in understanding whether the simulation of specific surfaces recognized by the subjects was satisfactory.

As far as the quality judgment is concerned, the data were based on all the answers different from "I don't know."

The mean of realism is higher for condition 2 compared to condition 1 for each surface with the exception of beach sand (which is almost equal). The analysis by means of a t-test reveals that differences are statistically significant for snow ($p = 0.0105$) and forest underbrush ($p < 0.01$). Analogously, the mean of realism is higher for condition 2 compared to condition 3 for each surface with the exception of beach sand (which is almost equal). In particular, the differences are statistically significant for beach sand ($p < 0.01$) and snow ($p < 0.01$). Moreover, the mean of realism is higher for condition 1 compared to condition 3 for each surface with the exception of forest underbrush, which is minor. Differences are statistically significant for beach sand ($p < 0.01$), and for forest underbrush ($p = 0.0344$), which, as said, is greater for experiment 3.

As regards the mean of quality, it is higher for condition 2 compared to condition 1, with statistically significant differences for all the surfaces with the exception of dry leaves: beach sand ($p < 0.01$), gravel ($p = 0.0217$), snow ($p < 0.01$), and forest underbrush ($p = 0.0219$). The mean of quality is higher for condition 2 compared to condition 3 for each surface, and in particular the differences are statistically significant for beach sand ($p < 0.01$), for snow ($p < 0.01$). Furthermore, the mean of quality is similar for condition 1 compared to condition 3, with the exception of forest underbrush for which it is higher in condition 3 compared to condition 1, with statistically significant differences ($p = 0.032$).

The comparison about the percentages of "I don't know" answers reveals that for each surface they are higher for condition 1 compared to condition 2, and for condition 3 compared to condition 2. In addition, they are higher for condition 3 compared to condition 1, for each surface with the exception of forest underbrush (which is minor).

As regards the percentages of correct answers about the soundscapes presented, they are higher for condition 2 compared to condition 3, and in particular, the differences are statistically significant for the ski slope soundscape ($p < 0.01$).

Overall, subjects observed that soundscapes play an important role in recognition of the surfaces, precisely for their ability to create a context. Especially in terms of conflicting cues, as it was the case in condition 3, subjects were trying to identify the strongest cues, i.e., the element which had the strongest recognition factor. Sometimes the subjects found this task quite hard to complete, and this is why the percentage of "I don't know" answers is higher in condition 3 as opposed to condition 2.

When leaving a comment, several subjects observed that the simulation of snow was extremely realistic. This observation is also confirmed by the high degree of realism ($\mathrm{mean} = 5.3$) and quality ($\mathrm{mean} = 5.1$) with which the surface was rated.

On the other hand, for some subjects the concept of dry leaves was rather confusing, and this is also confirmed by the low recognition rate of that surface.

Overall, this experiment represents a strong indication of the importance of context in the recognition of a virtual auditory place, where self-sounds created by users' footsteps and soundscapes are combined. Further investigations are needed to enhance the realism of the simulated soundscape, in particular, by having the auditory cues changing according to the motion of the subject in the space.

6 Conclusions and Future Work

In this paper, we describe a system able to sonically simulate the act of walking on different virtual spaces. Two experiments are also presented, whose goal is to understand the ability of subjects to recognize the surfaces they were exposed to, and the role of soundscape design in creating a sense of place and context when implementing a virtual auditory walking experience.

In the second experiment described, the user was not able to interact with the soundscapes, which were created using precomposed soundtracks. This slightly reduces the possibility of reproducing a natural experience. As an example, in the real world subjects can affect the surrounding soundscape, for example, by hitting sound producing objects with their feet while walking. On the other hand, subjects could interact with the physically simulated sounds of footsteps generated by their own act of walking.

Results indicate the ability of subjects to recognize the virtually simulated surfaces, together with the importance of the addition of environmental sounds in enhancing the recognition. As shown in [31], results are consistent with those obtained when exposing subjects to real surfaces, which is an indication of the quality of the footsteps synthesizer. The experiments also stimulated some interesting observations regarding the interface adopted and the way the sound synthesis algorithms were controlled. As can be seen in Fig. 2, the board used is limited in size and does not allow a natural walking experience. In future investigations, we indeed decided to control the sound synthesis algorithms by using shoes enhanced with sensors [24]. This choice was motivated by the fact that we wanted to enhance the simulation with haptic feedback, provided by some actuators placed on the sole of the shoes. The use of haptic feedback on the shoes allows a realistic simulation of the tactile sensation of walking on different surfaces, since this is quite different in solid versus aggregate surfaces, as an example. The switch to shoes enhanced with sensors, however, also allowed a wider walking possibility for the subjects, now limited only by the actual size of the laboratory. By enhancing the simulation of auditory feedback with haptic feedback, experiments with both coherent and incoherent stimuli were performed, in order to understand the role of the auditory and haptic modality in recognizing the virtual surfaces subjects are exposed to. Results show that subjects perform better in the recognition task when exposed to auditory feedback rather than haptic feedback, and the combination of auditory and haptic feedback does not significantly enhance the recognition [24], [37].

Overall, the results described are an interesting starting point for further investigations on the role of self-sounds and environmental sounds to create a sense of place in a virtual environment. While walking an acting in an environment, a person is exposed to her own self-sounds as well as the sounds of the place. This paper presents a preliminary investigation of the role of these different elements both taken in isolation and combined. Further investigations are needed to gain a better understanding of the cognitive factors involved when subjects are exposed to different sound events, especially when a situation of semantic incongruence is present.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community' s Seventh Framework Programme under FET-Open grant agreement 222107 NIW—Natural Interactive Walking (www.niwproject.eu).

REFERENCES

[1] J. Adrien, *The Missing Link: Modal Synthesis.* MIT Press, 1991.

[2] F. Avanzini and D. Rocchesso, "Modeling Collision Sounds: Nonlinear Contact Force," *Proc. COST-G6 Conf. Digital Audio Effects (DAFx-01),* pp. 61-66, 2001.

[3] A. Benbasat, S. Morris, and J. Paradiso, "A Wireless Modular Sensor Architecture and Its Application in On-Shoe Gait Analysis," *Proc. IEEE Sensors,* vol. 2, 2003.

[4] B. Blesser and L. Salter, *Spaces Speak, Are You Listening?: Experiencing Aural Architecture.* MIT Press, 2006.

[5] M. Chion and C. Gorbman, *Audio-Vision,* Nathan, 1990.

[6] P. Chueng and P. Marsden, "Designing Auditory Spaces to Support Sense of Place: The Role of Expectation," *Proc. Computer Supported Cooperative Work Workshop: The Role of Place in Shaping Virtual Community,* 2002.

[7] P. Cook, "Physically Informed Sonic Modeling (PhISM): Synthesis of Percussive Sounds," *Computer Music J.,* vol. 21, no. 3, pp. 38-49, 1997.

[8] P. Cook, "Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds," *Proc. 22nd Int'l Conf. Virtual, Synthetic, and Entertainment Audio,* pp. 73-78, 2002.

[9] P. Cook, *Real Sound Synthesis for Interactive Applications.* AK Peters, Ltd., 2002.

[10] P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter, "Single State Elastoplastic Friction Models," *IEEE Trans. Automatic Control,* vol. 47, no. 5, pp. 787-792, May 2002.

[11] A. Farnell, "Marching Onwards: Procedural Synthetic Footsteps for Video Games and Animation," *Proc. Pure Data Convention,* 2007.

[12] F. Fontana and R. Bresin, "Physics-Based Sound Synthesis and Control: Crushing, Walking and Running by Crumpling Sounds," *Proc. Colloquium Musical Informatics,* pp. 109-114, May 2003.

[13] J. Freeman and J. Lessiter, "Hear There & Everywhere: the Effects of Multi-channel Audio on Presence," *Proc. Int'l Conf. Auditory Display (ICAD '01),* pp. 231-234, 2001.

[14] W. Gaver, "What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception," *Ecological Psychology,* vol. 5, no. 1, pp. 1-29, 1993.

[15] B. Giordano and S. Mcadams, "Material Identification of Real Impact Sounds: Effects of Size Variation in Steel, Glass, Wood, and Plexiglass Plates," *J. Acoustical Soc. Am.,* vol. 119, pp. 1171-1181, 2006.

[16] B. Giordano, S. Mcadams, Y. Visell, J. Cooperstock, H. Yao, and V. Hayward, "Nonvisual Identification of Walking Grounds," *J. Acoustical Soc. Am.,* vol. 123, no. 5, p. 3412, 2008.

[17] K.H. Hunt and F.R.E. Crossley, "Coefficient of Restitution Interpreted as Damping in Vibroimpact," *J. Applied Mechanics,* vol. 42, no. 2, pp. 440-445, 1975.

[18] B. Kapralos, D. Zikovitz, M. Jenkin, and L. Harris, "Auditory Cues in the Perception of Self-Motion," *Proc. 116th AES Convention,* 2004.

[19] R. Klatzky, D. Pai, and E. Krotkov, "Perception of Material from Contact Sounds," *Presence: Teleoperators and Virtual Environments,* vol. 9, no. 4, pp. 399-410, 2000.

[20] P. Larsson, D. Västfjäll, and M. Kleiner, "Perception of Self-Motion and Presence in Auditory Virtual Environments," *Proc. Seventh Ann. Workshop Presence,* pp. 252-258, 2004.

[21] A. Law, B. Peck, Y. Visell, P. Kry, and J. Cooperstock, "A Multi-Modal Floor-Space for Experiencing Material Deformation Underfoot in Virtual Reality," *Proc. IEEE Int'l Workshop Haptic Audio Visual Environments (HAVE '08) and Games,* pp. 126-131, 2008.

[22] X. Li, R. Logan, and R. Pastore, "Perception of Acoustic Source Characteristics: Walking Sounds," *J. Acoustical Soc. Am.,* vol. 90, pp. 3036-3049, 1991.

[23] R. Nordahl, "Increasing the Motion of Users in Photorealistic Virtual Environments by Utilizing Auditory Rendering of the Environment and Ego-Motion," *Proc. Presence,* pp. 57-62, 2006.

[24] R. Nordahl, A. Berrezag, S. Dimitrov, L. Turchet, V. Hayward, and S. Serafin, "Preliminary Experiment Combining Virtual Reality Haptic Shoes and Audio Synthesis," *Proc. Int'l Conf. Haptics: Generating and Perceiving Tangible Sensations,* pp. 123-129, 2010.

[25] R. Nordahl, S. Serafin, and L. Turchet, "Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications," *Proc. IEEE Virtual Reality Conf.,* 2010.

[26] S. O'Modhrain and G. Essl, "Pebblebox and Crumblebag: Tactile Interfaces for Granular Synthesis," *Proc. Conf. New Interfaces for Musical Expression (NIME '04),* pp. 74-79, 2004.

[27] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* second ed. McGraw-Hill, 1984.

[28] J. Paradiso, K. Hsiao, and E. Hu, "Interactive Music for Instrumented Dancing Shoes," *Proc. Int'l Computer Music Conf.,* pp. 453-456, 1999.

[29] R. Pastore, J. Flint, J. Gaston, and M. Solomon, "Auditory Event Perception: The Source-Perception Loop for Posture in Human Gait," *Perception and Psychophysics,* vol. 70, no. 1, pp. 13-29, 2008.

[30] L. Peltola, C. Erkut, P. Cook, and V. Valimaki, "Synthesis of Hand Clapping Sounds," *IEEE Trans. Audio, Speech, and Language Processing,* vol. 15, no. 3, pp. 1021-1029, Mar. 2007.

[31] S.S.R. Nordahl and L. Turchet, "Sound Synthesis and Evaluation of Interactive Footsteps for Virtual Reality Applications," *Proc. IEEE Virtual Reality Conf.,* 2010.

[32] N. Raghuvanshi and M. Lin, "Interactive Sound Synthesis for Large Scale Environments," *Proc. Symp. Interactive 3D Graphics and Games,* pp. 101-108, 2006.

[33] R. Sanders Jr., "The Effect of Sound Delivery Methods on a Users Sense of Presence in a Virtual Environment," PhD thesis, Naval Post Graduate School, 2002.

[34] R. Schafer, *The Tuning of the World.* Alfred A. Knopf, 1977.

[35] S. Serafin, L. Turchet, and R. Nordahl, "Extraction of Ground Reaction Forces for Real-Time Synthesis of Walking Sounds," *Proc. Audio Mostly Conf.,* 2009.

[36] R. Storms and M. Zyda, "Interactions in Perceived Quality of Auditory-Visual Displays," *Presence: Teleoperators and Virtual Environments,* vol. 9, no. 6, pp. 557-580, 2000.

[37] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Conflicting Audio-Haptic Feedback in Physically Based Simulation of Walking Sounds," *Proc. Int'l Conf. Haptic and Audio Interaction Design (HAID '10),* p. 97, Sept. 2010.

[38] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner, "Travelling without Moving: Auditory Scene Cues for Translational Self-Motion," *Proc. Int'l Conf. Auditory Display (ICAD '05),* 2005.

[39] K. Van Den Doel, P. Kry, and D. Pai, "FoleyAutomatic: Physically-Based Sound Effects for Interactive Simulation and Animation," *Proc. 28th Ann. Conf. Computer Graphics and Interactive Techniques,* pp. 537-544, 2001.

[40] D. Västfjäll, "The Subjective Sense of Presence, Emotion Recognition, and Experienced Emotions in Auditory Virtual Environments," *CyberPsychology & Behavior,* vol. 6, no. 2, pp. 181-188, 2003.

Rolf Nordahl is currently an assistant professor in Medialogy at Aalborg University Copenhagen. He is principal investigator for the EU funded project Natural Interactive Walking, and has earlier done seminal work on the EU-project BENOGO (the project was focused on HMD-based photo-realistic VR). Likewise, he is a recognized member of the expert panel under the Danish Evaluation Institute as well as being member of various steering committees. He frequently publishes both journal and conference papers and is also giving talks at international level. Lately, he was invited to run a special series of lectures at among other places Yale University (Connecticut). His research interests lie within VR, (Tele)-Presence, Sonic Interaction Design, developing new methods and evaluation-techniques for VR, and Presence and Games.

Luca Turchet received the Laurea degree (summa cum laude) in computer science from the University of Verona, in 2006. At the same time, he studied classical guitar and composition at the Music Conservatory E.F. Dall'Abaco of Verona receiving the diploma degrees, in 2007 and 2009, respectively (both summa cum laude). In 2006, he was a visiting scholar at the Music Technology Group of Barcelona, Spain, and in 2008 at the Cork Institute of Technology-School of Music in Cork, Ireland. In 2009, he started the PhD at the Medialogy Department of the Aalborg University Copenhagen working on the Natural Interactive Walking Project. His main research interests include interactive sound design, multimedia and multimodal systems, sound spatialization, and human-computer interaction. He is a student member of the IEEE.

Stefania Serafin received the PhD degree in computer-based music theory and acoustics from Stanford University, in 2004, and a Master in Acoustics, computer science and signal processing applied to music from Ircam (Paris), in 1997. She is an associate professor in sound modeling at Aalborg University Copenhagen. She has been a visiting professor at the University of Virginia (2003), and a visiting scholar at Stanford University (1999), Cambridge University (2002), and KTH Stockholm (2003). She is a principal investigator for the EU funded project Natural Interactive Walking, and Danish delegate for the EU COST Action on Sonic Interaction Design. Her main research interests include sound models for interactive systems and multimodal interfaces, and sonic interaction design.