Available online at www.sciencedirect.com

ScienceDirect

Int. J. Human-Computer Studies 67 (2009) 947–959

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

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*

Received 20 September 2008; received in revised form 1 June 2009; accepted 21 July 2009
Available online 6 August 2009

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.

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).

1071-5819/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2009.07.007

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 $\mathbf{F}(\omega, t)$,

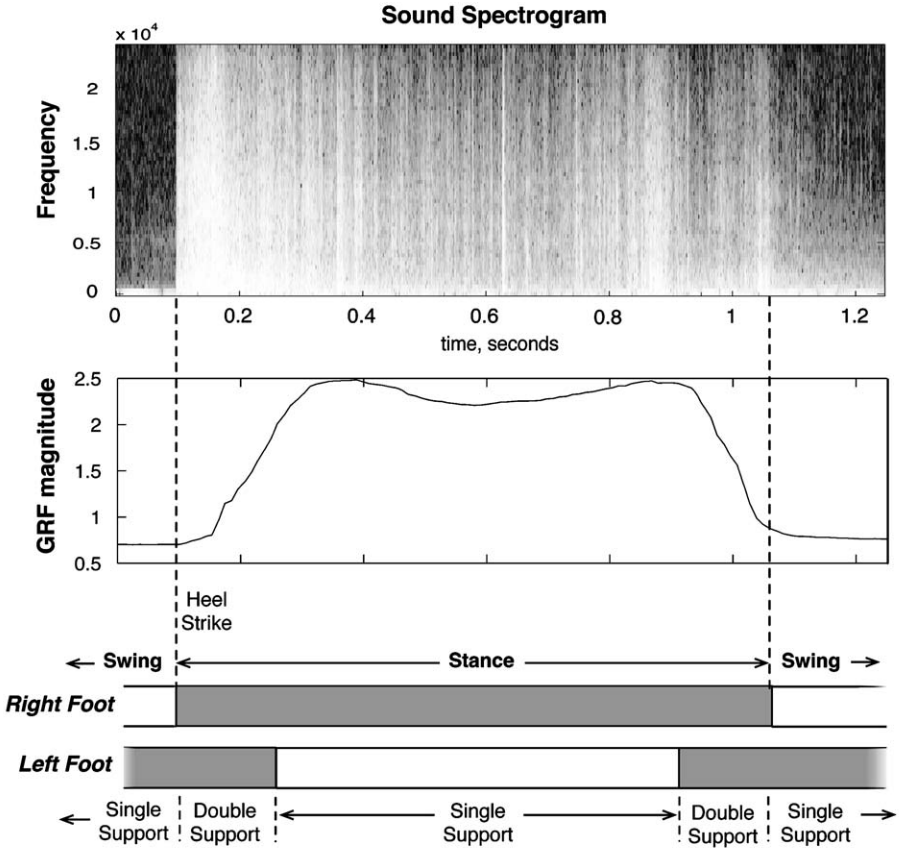

Fig. 1. (Left) Bottom: gait phase for left and right foot, showing more than 50% of the gait cycle. Middle: GRF for right foot (authors' recording). Top: spectrogram showing the acoustic signature resulting from a step onto gravel.

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

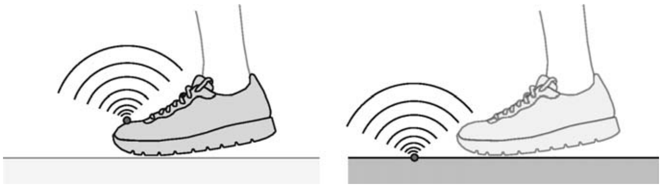

Fig. 2. A walking surface augmented using an instrumented shoe (left) or with an instrumented floor (right).

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

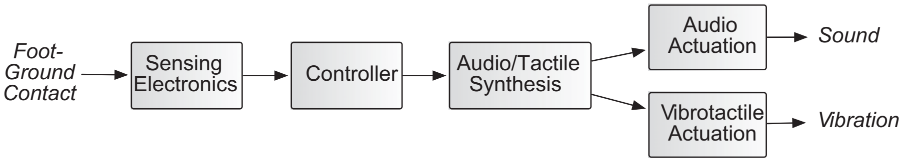

Fig. 3. Footstep interaction viewed as an open-loop process.

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.

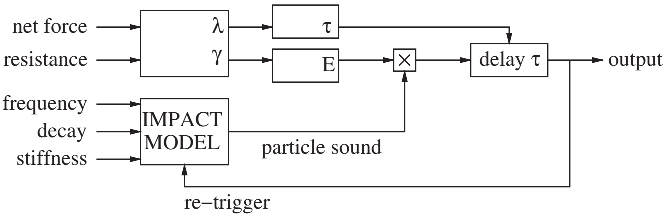

Fig. 4. Algorithm for the generation of sounds of contact with aggregate materials.

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.

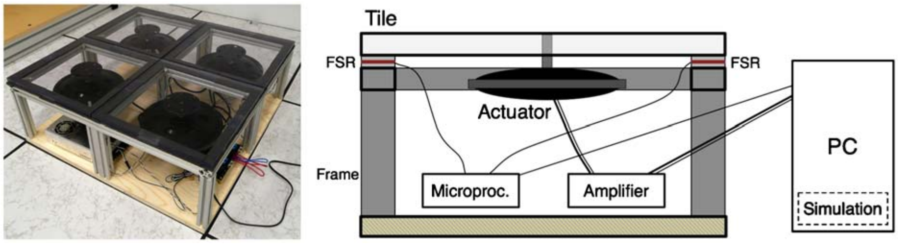

Fig. 5. Left: an image of the tile prototype, showing the tile surface (polycarbonate sheet), vibrotactile actuator, force-sensing resistors, structural frame, and associated electronics. Right: diagram of the same, including the PC running the floor material simulation.

prototype shown (Fig. 5) consists of a set of rigid 34 × 34 × 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

$$\lambda(t) = Au(t)(1 + \tanh(Bu))/2, \tag{1}$$

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

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

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

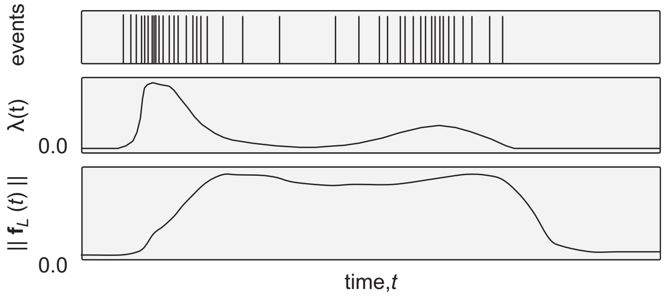

Fig. 6. Qualitative illustration of the control model of Eq. (2), relating the time derivative of the low frequency force signals $d\mathbf{f}_L/dt(t)$ (Bottom) to the event rate parameter $\lambda(t)$ (Middle) and a sampled event sequence (Top).

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

---

[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.

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

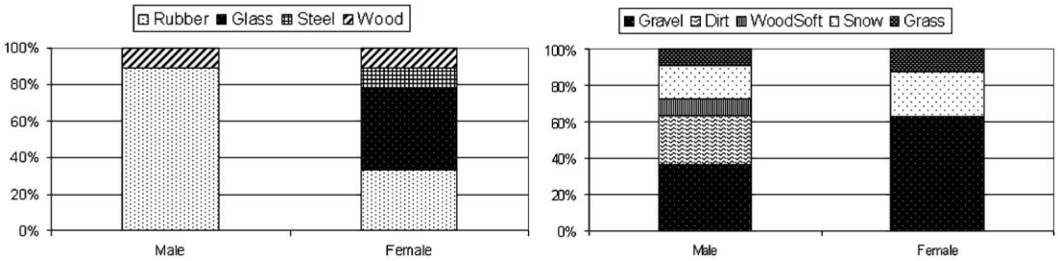

Fig. 7. Subjects' percentage choices of different ground materials in association to walker's gender. The left figure shows the choices for impact sound models. The right figure shows subjects' preferences for different tunings of the crumpling sound model.

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 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).

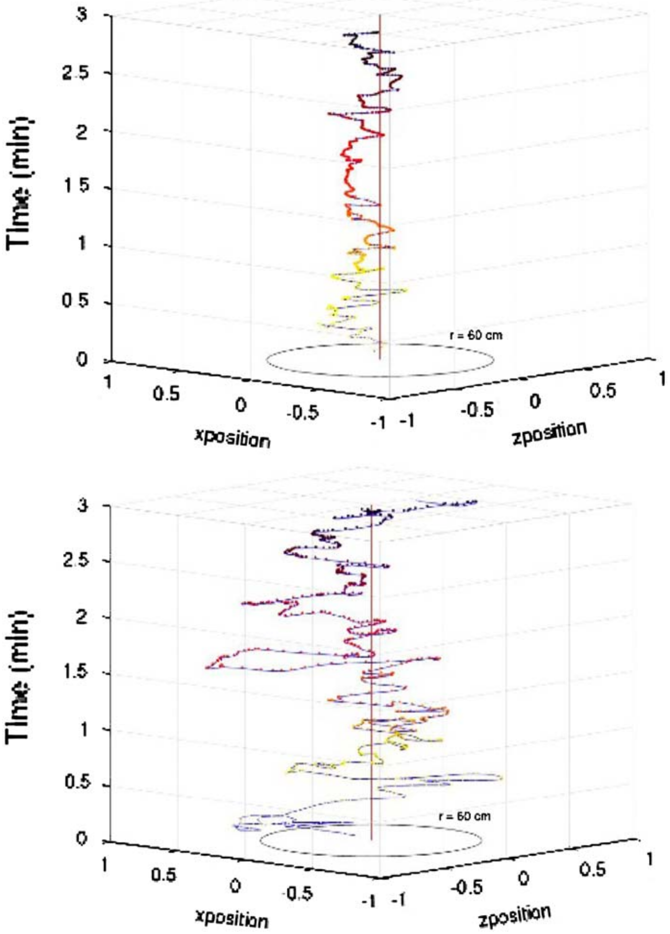

Fig. 9. Top: visualization over time of the motion of one subject over time with visual only condition (top) and full condition (bottom).

(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.

Acknowledgments

The authors gratefully acknowledge the support from the following sources: The European Commission for projects CLOSED (FP6-NEST-Path no. 029085), BrainTuning (FP6-NEST-Path no. 028570), and NIW (FP7-FET-Open no. 222107); The Quebec MDEIE for the project NIW; The ESF COST office for Actions no. 287 on Gesture Controlled Audio Systems (ConGAS) and no. IC0601 on Sonic Interaction Design (SID).

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.