Applied Acoustics 83 (2014) 22–31

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Custom made wireless systems for interactive footstep sounds synthesis

Luca Turchet *

Department of Architecture, Design and Media Technology, Aalborg University Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 30 August 2013
Received in revised form 6 March 2014
Accepted 10 March 2014
Available online 12 April 2014

Keywords:
Interactive sonification
Locomotion interface
Physical modeling

ABSTRACT

This paper describes both the hardware and software development of three custom made wireless systems used for the interactive synthesis of footstep sounds. The data collected from the detection of the feet movements of a walker are used for real-time control of physical models for the auditory display of different ground textures and shoe types. The first system is based on a wooden plank under which an array of microphones is placed. The second system exploits the motion capture technology. The third system consists of a pair of sandals enhanced with two force sensitive resistors and two 3-axes accelerometers for each shoe. The characteristics of the three architectures are discussed and compared. The developed locomotion interfaces find application in several contexts, such as augmented reality, virtual reality, or entertainment, as well as in perceptual studies investigating the influence of interactive sounds on locomotion performance usable for training and rehabilitation purposes.

1. Introduction

Sonic interaction design is a research area which explores methods to convey information (also in terms of aesthetic and emotional qualities) by means of sounds in interactive contexts [1,2]. On the one hand, this research field addresses the challenges of creating interactions mediated by the sound by means of designing and implementing novel interfaces to control sonic events as response to the gestures of one or more users. On the other hand, it investigates the action-perception loop deriving from the interaction with the developed interfaces. The user actively manipulating the designed sonic interfaces discovers how his/her actions modulate the sound. In addition, the auditory feedback can in turn guide the actions of the user by providing the information about how to modify the actions themselves.

Sonic interaction design is closely connected with a subtopic of the human–computer interaction field termed sonification [3]. Such a research field addresses how information can be conveyed in an auditory, typically non-speech, form. Basically, data of various nature are transformed into sound so that the listener is supported to better understand and interpret them. The study of the human interaction with a system that transforms data into sound is a subfield of sonification called interactive sonification [2].

A related concept important to the design of interfaces for human–computer interaction, and particularly significant for sound-based communication purposes, is that of embodied interaction. It has been defined by Dourish as "*the creation, manipulation and sharing of meaning through engaged interaction with artefacts*" [4]. According to Dourish meanings are necessarily present in the actions that people accomplish during the interaction with objects, with other people, and with the environment. As a consequence, perception and action are linked. Embodied interactions occur in real-time and real-space as a part of the world in which we are situated. As a result of this view, the so-called embodied interfaces allow for direct manipulation and are based on a closed loop paradigm where the control of the interfaces is exerted by the user via a continuous and simultaneous set of gestures and perceptions.

A specific case of interactive sonification regards the interactive transformation of actions into sounds. Object of the present work is the interactive sonification of a walker's feet movements tracked by different types of wireless locomotion interfaces. Lately, the interest towards the development of locomotion interfaces capable to provide real-time auditory feedback to the walker is noticeably growing [5]. This is also due to the wide range of scenarios in which such interfaces find application. In virtual reality contexts, augmented floors [6,7] have been developed to provide the user with footstep sounds while physically navigating in the virtual environments. In training and rehabilitation contexts, instrumented insole systems have been built to present corrective information to the patient by means of auditory feedback [8,9]. In interactive art, especially dance, shoes enhanced with sensors have also been developed [10,11]. On a different vein, the attention of researchers has been devoted to the creation of shoes for real-time gait analysis [12–14]. Requirements in the various situations are

* Tel.: +45 9940 2498.
E-mail address: tur@create.aau.dk

http://dx.doi.org/10.1016/j.apacoust.2014.03.005
0003-682X/

obviously quite different. While in gait analysis research accuracy of the system is the most important factor, in interactive arts and virtual reality applications the main design parameter is timeliness: a system with a considerable latency is evaluated as unusable in any real-time multimedia application.

On the other hand, most of the research efforts on the synthesis of footstep sounds have been focused on algorithmic solutions not suitable for a direct parametric control during the act of walking [15–18]. A noticeable exception is the sound synthesis engine proposed in [7], capable to interactively sonify foot movements into footstep sounds on various types of surface materials. Recently that work has been extended, allowing the simulation of various types of shoes and the modeling of some anthropomorphic features such as gender and weight of the walker [19]. Such an engine was based on physical models which were driven by an exciter signal expressing the type of foot–floor interaction (e.g., walking, running, sliding, jumping). Various systems for the generation of such an input in real-time with the walker's foot movements were developed and tested [7,20–24].

One of these systems consisted of microphones, placed on the floor at the corners of a square configuration, that detected the footstep sounds generated by a user [7]. Subsequently, the captured signal was provided as input to the synthesis engine, in order to simulate footstep sounds on materials different from the one the user was walking on. This apparatus allowed to reach the requisite of shoe independence and it was very accurate in the detection of the user's feet movements. However, the user was required to navigate in a specific location delimited by the space inside the microphones. Furthermore, this method required that the environment was quiet, since the microphones had to capture only the user footstep sounds, and any other signal constituted a not negligible input error for the footstep synthesizer. This means that the sound could not be delivered to the user through loudspeakers but it had to be conveyed through headphones.

A similar architecture based on a set of accelerometers was also proposed in [7]. The accelerometers were attached to a board where the users could walk upon, with the goal of capturing the signal propagated through the board and thus obtaining an audio range signal expressing the foot–surface interaction. This system presented the same limitations of the microphones-based system with in addition the problem of a lower level of accuracy in the detection of the footsteps dynamics.

Another system that was proposed consisted of shoes enhanced with pressure sensors, which triggered the footstep sounds according to the steps of a user [20]. Such an approach was not shoe-independent and made use of wires connected to the shoes, with the disadvantage of preventing a completely free navigation. More importantly, it did not take into account the exact step movement made by the user, therefore the resulting auditory feedback suffered a lack of realism. In addition, it was not possible to detect when the user was sliding the foot on the floor since the pressure sensors alone were not enough to serve this purpose. Such a system was preliminarily evaluated with user experiments reported in [24]. Results showed that users judged the interaction with the system not too much natural and that they felt quite constrained during the act of walking. Users reported that the main cause for these results was the presence of the wires.

Starting from the architectures proposed in previous research, as well as from the experimental results achieved with the testing of those solutions, three novel wireless interfaces have been developed to advance the state of the art in the research on interactive sonification of a walker's feet movements. The focus in the design of such interface was to provide the walker with stimuli valid from the ecological point of view [25–27]. To achieve such a goal, three requirements were set in the design of the three prototypes: (i) real-time control of the footstep sounds synthesizer; (ii) accuracy of the feet movements detection in order to achieve a wide range of dynamics in the produced sounds; (iii) freedom of navigation when interacting with the systems; and (iv) embodiment of the interaction.

2. Synthesis of footstep sounds

The utilized sound synthesis technique was based on considering a footstep sound as the result of multiple micro-impact sounds between the shoe and the floor. The set of such micro-events was considered as a high level model of impact between an exciter (the shoe) and a resonator (the floor). The synthesis of a footstep sound on different kinds of materials was achieved starting from a signal in the audio domain containing a generic footstep sound on an arbitrary material. It consisted in removing the contribution of the resonator, keeping the exciter and considering the latter as input for a new resonator which implements different kinds of floors. Subsequently the contribution of the shoe and of the new floor were summed in order to have a complete footstep sound.

In [28] the problem of extrapolating the exciter from the acoustic waveform of a footstep sound was addressed. The results of such research led to the conclusion that an optimal solution to obtain such an exciter consisted of extracting its amplitude envelope. The envelope (e) was extracted from the signal (x) by means of the non-linear low-pass filter proposed in [29] and subsequently utilized in [15]:

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

where n and $n-1$ indicate respectively the current and previous sample (sample rate 44,100 Hz) of the discretized variable they refer to, and b_{up} and b_{down} are equal to 0.8 and 0.995 respectively. Fig. 1 shows both the waveform and the corresponding envelope of a typical footstep sound on a concrete floor as well as the sound resulting from the sliding of the foot on the same floor.

In order to simulate the footstep sounds on different types of materials, the envelope extracted with this technique was used to control various sound synthesis algorithms based on physical models. Specifically, the involved sound models were those described in [30,31] for impacts, in [32] for frictions, in [33] for crumpling events, in [15] for particles interactions (PhISM), and in [19] for solid–liquid interactions. By using such models either alone or in combination with each other, the simulation of a large palette of footstep sounds on solid (e.g., wood), liquid (e.g., puddles), and aggregate surfaces (e.g., gravel) was achieved.

On the other hand, various types of shoes were simulated by using different types of exciter signals along with an appropriate tuning of the parameters controlling the involved sound models. For a detailed description of the above mentioned approaches the reader is referred to [7,19]. The proposed footsteps synthesizer was implemented in the Max/MSP sound synthesis and multimedia real-time platform.

3. Microphones-based system

This section describes a non-intrusive shoe-independent system conceived as evolution of the architecture involving four microphones mentioned in Section 1. Such a system consisted of a wooden plank (5.40 m long, 90 cm wide and 7 cm tall, under which a set of nine microphones was placed to detect the feet–floor interaction (see Fig. 2(b)). The microphones (Line Audio Design CM3) were arranged on the floor at an equal distance of 60 cm from each other along the line lying at the half of the plank width. In particular, each microphone was attached to the floor by

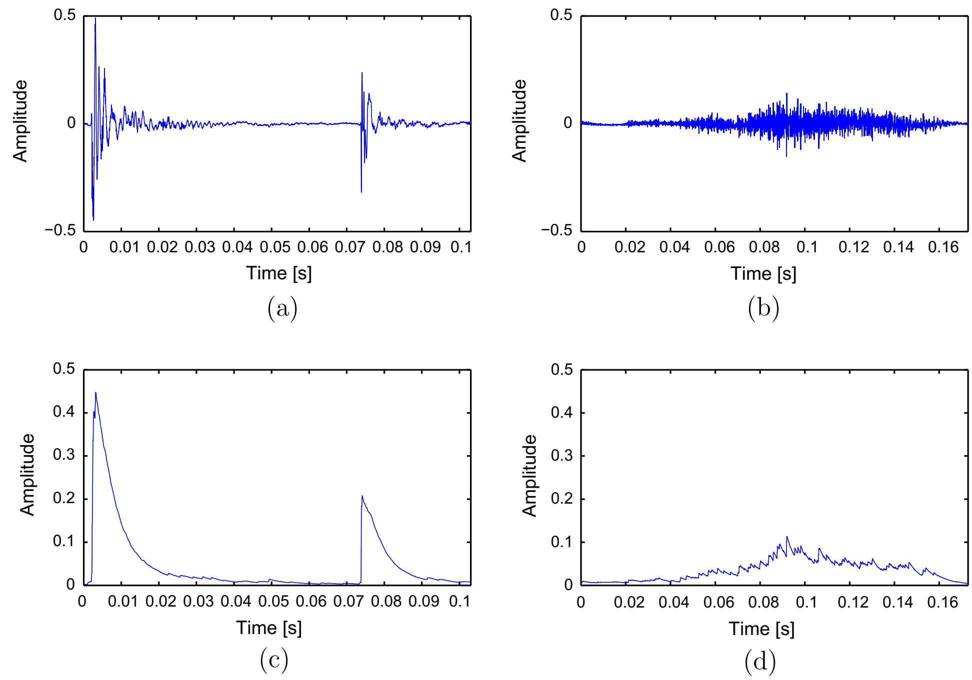

Fig. 1. Waveforms (top) and relative extracted envelope (bottom) of a footstep sound on a concrete floor (left) and of the sound resulting from the sliding of the foot on the same floor (right). In the envelope of the footstep sound (Fig. 1(c)) it is possible to notice the sub-events heel and toe.

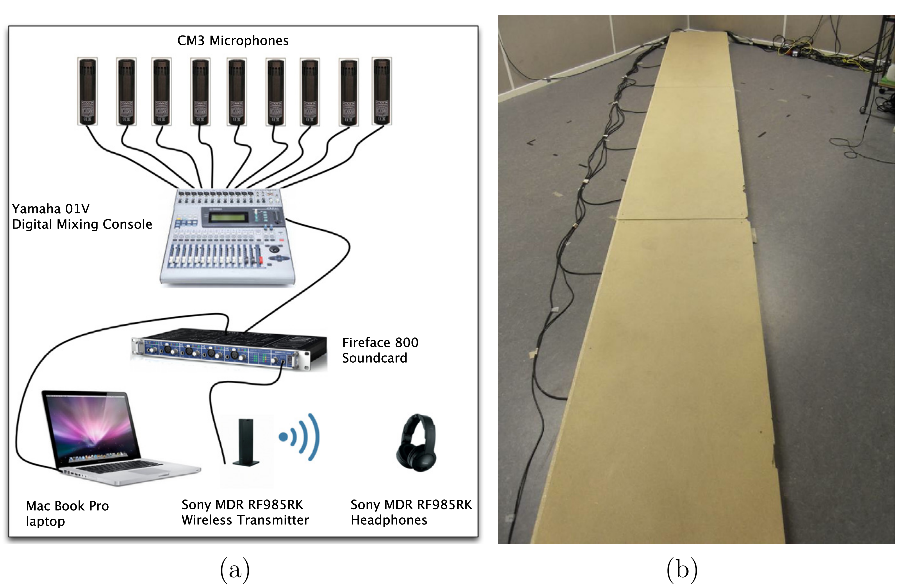

Fig. 2. The block diagram of the microphones-based system (Fig. 2(a)), and the pictures of the wooden plank (Fig. 2(b)) under which the array of microphones was placed.

means of scotch tape, and was lifted of 2 cm from the floor thanks to a layer of foam. The microphones signals were mixed together (through a Yamaha 01 V Digital Mixing Console), and subsequently digitalised by means of a soundcard (Fireface 800) connected to a laptop running the footstep sounds synthesizer, as illustrated in Fig. 2(a). The real footstep sounds produced by a walker and detected by the microphones were used to control the temporal evolution of the synthetic footsteps according to the real-time synthesis techniques developed in previous research (see Section 2). Before performing the envelope extraction, the incoming signal was processed with an algorithm of noise reduction. For this purpose the real-time noise reduction tools available in the FFTease collection of externals for Max/MSP were utilized [34]. The synthesized sounds were finally conveyed to the walker by means of a closed analog wireless headphone set (Sony MDR RF985RK). Fig. 3 illustrates the waveforms of the signal detected by the set of microphones and of the corresponding synthesized footstep sound on gravel.

The placement of the microphones under the wooden plank instead of outside it, as done in previous research [7], allowed to track exclusively the interaction between the feet and the floor, since this approach avoided the detection of other sound sources (such as the rubbing of pants leg material, the squeak of the shoes, the noise produced by the air movement on the microphone due to fast movements, or even voices and coughing) which constituted unwanted input signals for the sound synthesis engine. Moreover,

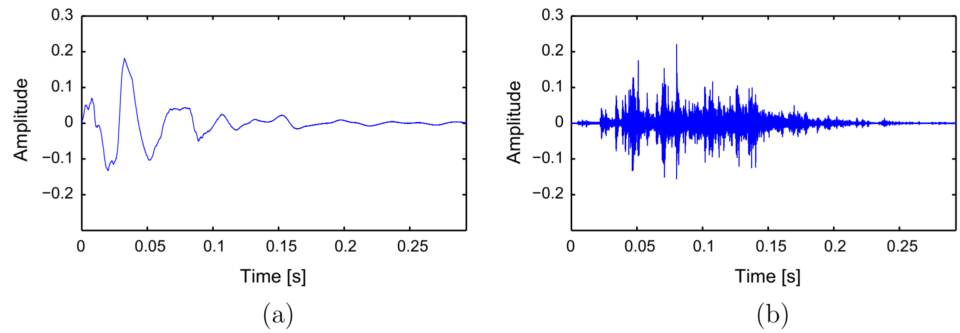

Fig. 3. Waveforms of the signal detected by the set of microphones (Fig. 3(a)) during a step using high heel on the wooden plank, and of the corresponding synthesized footstep sound on gravel (Fig. 3(b)).

particular care was given to the accomplishment of an identical detection accuracy in each point of the plank, in order to achieve a correct mapping in terms of dynamics between the walkers steps and the corresponding sounds interactively produced.

4. MoCap-based system

This section addresses the problem of detecting a walker's feet movements from the data tracked by means of a Motion Capture System (MoCap) in order to provide a real-time control of the footsteps synthesizer.

Fig. 4 shows a schematic representation of the overall architecture developed. This system was placed in an acoustically isolated laboratory and consisted of a MoCap (Optitrack), two soundcards (FireFace 800), sixteen loudspeakers (Dynaudio BM5A), and two computers. The first computer run the motion capture software (Tracking Tools 2.0), while the second run the audio synthesis engine. The two computers were connected through an ethernet cable and communicated by means of the UDP protocol. The data relative to the MoCap were sent from the first to the second computer which processed them in order to control the sound engine. The MoCap was composed by sixteen infrared cameras (OptiTrack FLEX:V100R2) which were placed in a configuration optimized for the tracking of the feet. In order to achieve this goal, two sets of markers were placed on each shoe worn by the subjects, in correspondence to the heel and to the toe respectively. As far as the auditory feedback is concerned, the sounds were delivered through a set of sixteen loudspeakers placed on the ground.

The data coming from the four sets of markers were processed in order to generate interactively the exciter signal to control the footstep sounds synthesizer. The developed algorithms were based on the triggering of recordings containing different typologies of exciters (those corresponding to a step and to the sliding the foot on the floor). During the user's locomotion a variation of the values of the markers coordinates occurred in correspondence to each step. In particular, the z coordinate (see Fig. 4) was taken into account when detecting whether the foot was on air or on the ground, and when it hit the ground. Each time the value of the z coordinate of one of the sets of markers passed from being on air to being on the ground, the exciter corresponding to that set of markers (i.e., heel or toe) was triggered into the footstep sounds engine (see Fig. 5). More precisely, only negative changes in the z coordinate were checked, since the focus was in the generation of the sound when the step hit the ground, and not when leaving it. This passage was detected by means of a set of thresholds. In addition, the amplitude of each triggered exciter was proportionally controlled by the absolute value of the velocity along the z axis, being this parameter related to the intensity with which the foot hit the ground.

Other thresholds were used in order to handle some boundary conditions, like the standing of the subject, with the aim of controlling the sound generation. Such thresholds were set in a phase of calibration of the system.

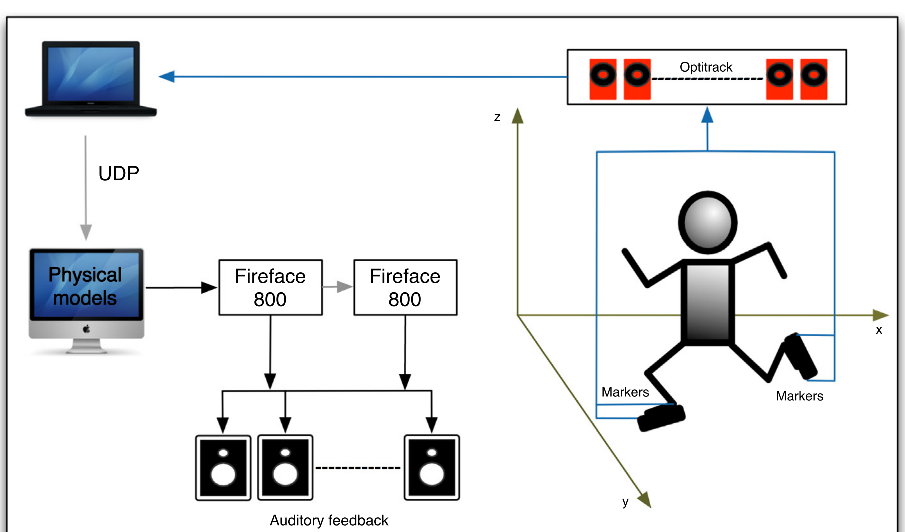

Fig. 4. A block diagram of the MoCap-based system and the used reference coordinates system.

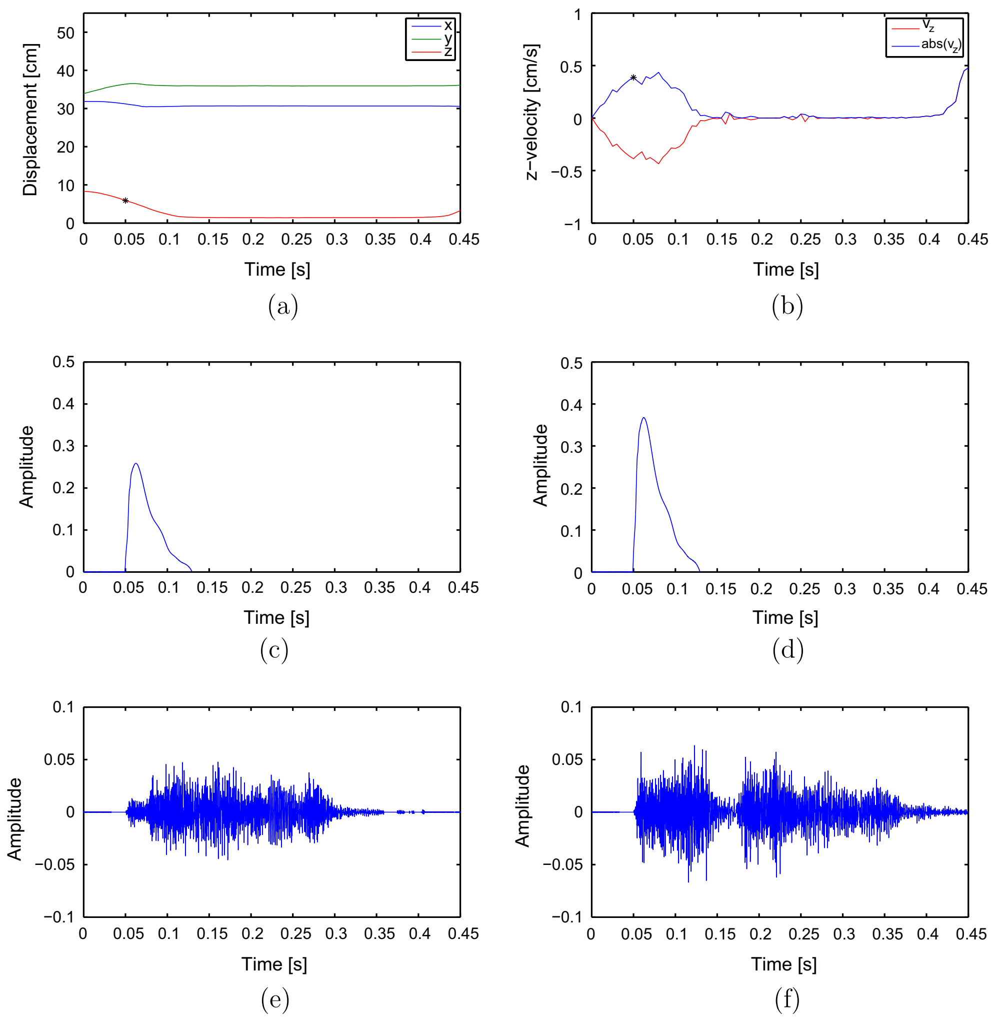

Fig. 5. Synthesis process using a marker placed on the heel tracked by a MoCap system: from the sensor data corresponding to the heel to the simulated sound on a water puddle wearing sneakers. Fig. 5(a) shows the temporal evolution of the marker displacement along the three axis; the black asterisk indicates the moment of triggering of the exciter, i.e., when the marker assumes *z*-values smaller than a threshold; Fig. 5(b) shows the corresponding velocity along the *z* axis; the black asterisk indicates the *z*-velocity absolute value detected at the moment of triggering of the exciter. Fig. 5(c) shows the exciter utilized for simulating the heel of sneakers shoes, while Fig. 5(d) illustrates the same exciter modulated in amplitude by the *z*-velocity absolute value. Such a difference in amplitude is reflected in the corresponding synthesized sound as shown in Fig. 5(e and f).

The system was also able to track the sliding of the foot on the ground. Such action produced the triggering of an exciter signal whose envelope was modulated by a linear combination of the absolute value of the velocity of the foot along the *x* and *y* coordinates. For this purpose it was possible to consider only the data coming from the set of markers placed on the heel.

The triggered exciters were created ad hoc by means of MATLAB procedures, as described in [19]. For each simulated shoe type (e.g., high heels, dress shoes, boots, sneakers), five types of heel and toes signals were used and randomly chosen at the moment of the triggering, giving rise to twenty-five possible combinations. Similarly, three exciters were used for the sliding case. Such behavior was adopted in order not to have always the same type of exciter as input of the engine, and this allowed to have differences in the generated sounds at every step, increasing thus the degree of realism of the walking experience.

5. Wireless sandals

The setup for the developed shoe-integrated sensors system is illustrated in Fig. 6. Such a system was composed by a laptop, a wireless data acquisition system (DAQ), and a pair of sandals each of which was equipped with two force sensing resistors (FSR) (Interlink 402) and two 3-axes accelerometers (ADXL325). In more detail, the two FSR sensors were placed under the insole in correspondence to the heel and toe respectively. Their aim was to detect the pressure force of the feet during the locomotion of the walker. The two accelerometers instead were fixed inside the sandals. Two cavities were made in the thickness of the sole to accommodate them in correspondence to the heel and toe respectively. In order to better fix the accelerometers to the sandals the two cavities containing them were filled with glue.

The analogue sensor values were transmitted to the laptop by means of a portable and wearable DAQ illustrated in Fig. 7. The wireless DAQ consisted of three boards: an Arduino MEGA 2560 board, a custom analog preamplifier board, and a Watterott RedFly wireless shield. In the nomenclature of the Arduino community, a "shield" is a printed circuit board that matches the layout of the I/O pins on a given Arduino board, allowing that a shield can be stacked onto the Arduino board, with stable mechanical and electrical connections. As shown in Fig. 7, all three boards were stacked together. In this way the wireless DAQ system could be easily put together in a single box, to which a battery was attached. This

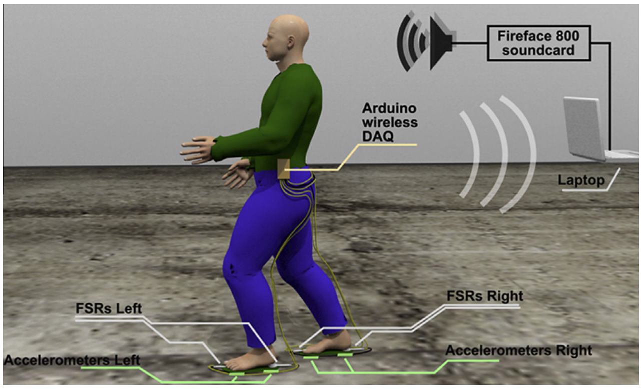

Fig. 6. Setup for the wireless sandals system: the user wears the sensor enhanced sandals and the wireless data acquisition system.

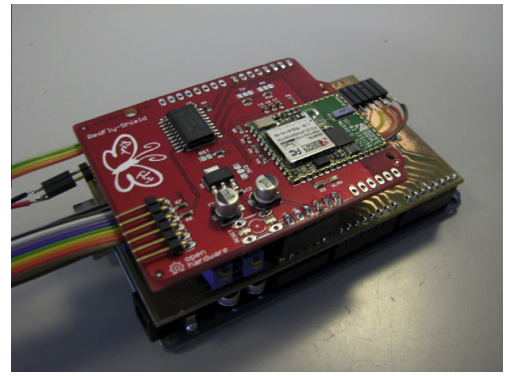

Fig. 7. The wireless data acquisition system. It was composed of a stack of Arduino MEGA 2560 (bottom) and of an additional circuitry which consisted of an analog preamplifier board (center) and the Watterott RedFly wireless shield (top).

resulted in a standalone, portable device that could be attached to the user's clothes, allowing greater freedom of movement for the user.

Since each foot carried two FSRs and two 3-axes accelerometers, which together provided 8 analog channels of data, the system demanded capability to process 16 analog inputs in total. That was precisely the number of analog inputs offered by an Arduino MEGA 2560, whose role was to sample the input channels, and format and pass the data to the wireless RedFly shield. The analog preamplifier board was a collection of four quad rail-to-rail operational amplifier chips (STmicroelectronics TS924), providing 16 voltage followers for input buffering of the 16 analog signals, of four trimmers, to complete the voltage divider of the FSR sensors, and of connectors. The Watterott RedFly shield was based on a Redpine Signals RS9110-N-11–22 WLAN interface chipset, and communicated with the Arduino through serial (UART) at 230400 baud. Preliminary measurements showed that the entire wireless DAQ stack consumed about 200 mA with a 9 V power supply, therefore a power supply of 9 V as the battery format was chosen. The wireless shield acted as a UDP client, which streamed data to a Perl UDP bridge on the laptop, with a nominal sampling frequency of 418 Hz (for all channels) and latency of about 12 ms.

The data coming from both the pressure sensors (see Fig. 8(a)) and from the accelerometers (see Fig. 8(c)) were processed in the laptop in order to generate interactively the exciters for the control of the footstep sounds synthesizer. Analogously to the case of the MoCap based system, the developed algorithms were based on the triggering of exciters signals corresponding to a step (heel and toe) and to the sliding the foot on the floor, whose envelope and evolution in time were controlled by the dynamics of the feet movements.

As regards the case of the steps, the exciters corresponding to heel and toe strikes were triggered according to the activation of the relative FSR sensors. In particular, the triggering occurred when the first derivative of the pressure sensor values became bigger than a threshold (see Fig. 8(b)). In addition, in order to render the information about the intensity with which the foot hit the ground, the amplitudes of the exciters corresponding to heel and toe were modulated according to the values of the corresponding accelerometers. For this purpose the sum of the absolute value of the accelerometers along the three axes (also known as L1-norm) was considered. Specifically, the maximum value of the L1-norm (see Fig. 8(d)) detected between two subsequent activations of the corresponding pressure sensor was used to modulate the amplitude of a predefined exciter (see Fig. 8(e and f)). As it is possible to notice in Fig. 8(d) such a value occurs immediately before the beginning of the pressure sensor activation during a step, that is just before the heel or the toe hits the ground.

As far as the sliding is concerned, such a type of foot–floor interaction was rendered by triggering the corresponding exciter and modulating its envelope with a linear combination of the absolute values of the accelerometers components along the plane parallel to the ground. For this purpose it was important that the accelerometers were placed inside the thickness of the sole in a position perfectly parallel to the ground. Moreover, for the sliding detection it was enough to consider only the data coming from the accelerometer placed on the heel.

The algorithms handled some other boundary conditions, such as the standing of the user, with the aim of controlling the sound generation. The triggered exciters were the same used for the MoCap based system, and they were randomly chosen at the moment of the triggering with the purpose of increasing the realism of the interaction.

6. Discussion

The development of the above described systems was based on the challenge of interactively computing an exciter signal as input for the footstep sound synthesis engine presented in [7,19]. While in the microphones-based system such a signal was directly computed form the captured data (i.e., the acoustic waveform), in the other two systems such a computation was done in a four phases process: (i) detection of the type of foot–floor interaction (i.e, sliding or stepping); (ii) calculation of the triggering instant; (iii) calculation of the movement dynamics; and (iv) triggering of the exciter. The reason for adopting such a triggering method rather than directly computing the exciter from the acquired data was due to the fact that the synthesis engine required as input for the physical models a signal having a sample rate of 44,100 Hz, while the data coming from both the MoCap-based system and the wireless sandals arrived with a much lower rate (respectively 100 Hz and 1000 Hz). More importantly, the features of the data acquired from both the systems (see Figs. 5(a) and 8(a and c)) were not suitable to be mapped into a temporal evolution typical of the exciters needed as input for the synthesis engine (see Fig. 1(c and d).

In the reminder of this section the three developed architectures are discussed and compared in terms of affordances, accuracy, timeliness, wearability, ecology and embodiment of the interaction, as well as possible application contexts. From the reported analysis of the systems it is possible to notice that all the requirements set in the design phase (see Section 1) were achieved.

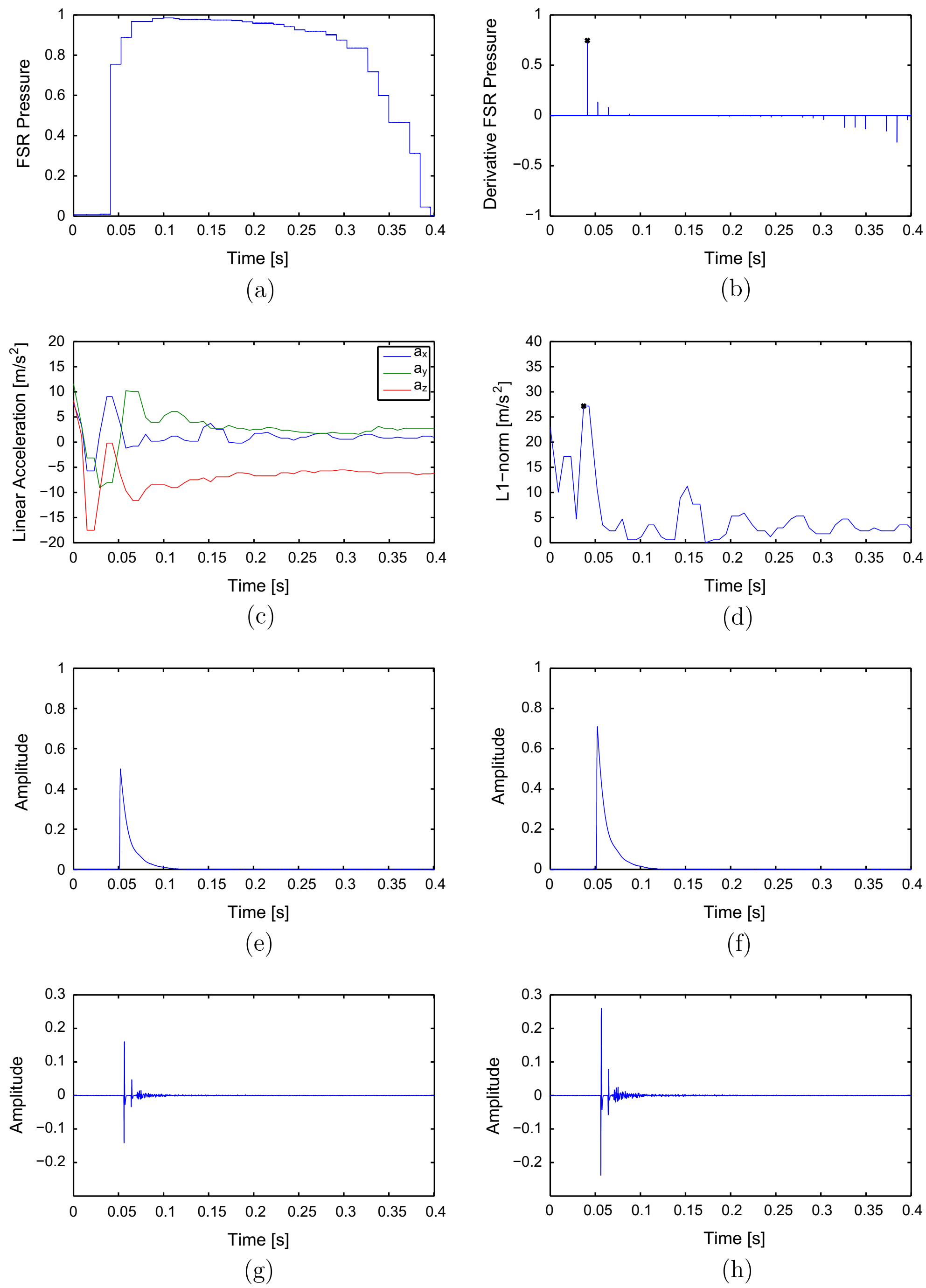

Fig. 8. Synthesis process using the wireless sandals: from the sensor data corresponding to the heel to the simulated sound on concrete wearing male dress shoes. Fig. 8(a) shows the temporal evolution of the pressure exerted by the heel during a step; Fig. 8(b) shows the corresponding first derivative; the black asterisk indicates the moment of triggering of the exciter, i.e. when the derivative becomes greater than a threshold. Fig. 8(c) illustrates the temporal evolution of the heel linear acceleration along the *x*, *y*, and *z* axes (adopting the same reference coordinate system of the MoCap system illustrated in Fig. 4). Fig. 8(d) shows the L1-norm of the accelerometers; the black asterisk indicates the maximum value detected immediately before the moment of triggering of the exciter. Fig. 8(e) shows the exciter utilized for simulating the heel of a male dress shoe, while Fig. 8(f) illustrates the same exciter modulated in amplitude by the maximum value of the L1-norm. Such a difference in amplitude is reflected in the corresponding synthesized sound as shown in Fig. 8(g and h).

6.1. Affordances

The three systems allow to interactively sonify several foot–floor interaction possibilities, including walking, walking in place, sliding, running, running in place, and jumping. In addition, the wireless sandals system allows the tracking and corresponding sonification of multiple users at the same time, something not possible with the other two systems: in the microphones-based system this is due to the inseparability of the signals corresponding to the two users, and in the MoCap-based system, to makers obscuration problems when moving. Moreover, two features which constitute a novelty compared to other shoes-integrated interfaces developed in previous research for the same purpose [20,11], are on the one hand the sliding detection, not possible without the use of 3-axes accelerometers, and on the other hand the mapping between the force with which the foot hits the floor and the amplitude of the corresponding synthesized sounds.

In a different vein, one of the salient differences between the microphones-based system and the other two is that the latter allow the separate tracking of the left and right foot, as well as of the heel and toe in each foot. As a consequence, this feature offers a range of sonification possibilities wider than that of the microphones-based system or other augmented floors currently developed [6]. For instance, it is possible to provide sounds different for the two feet or even different for heel and toe. In addition, differently from the microphones-based system, the MoCap-based system and the wireless sandals allow the simulation of types of shoes different from those worn by the user, due to the triggering of ad hoc built exciter signals [19]. Furthermore, the two systems allow a greater number of gait parameters tracked in real-time (e.g., feet pressure, step length, accelerations, trajectory) than those achievable from the analysis of the waveforms detected by the microphones (e.g., step time, step duration).

In all the three setups the user is free to navigate as no wires limiting the locomotion are involved. However, the walking area in the case of the MoCap-based system is delimited by the coverage angle of the infrared cameras, and in the case of the microphones-based system is delimited by the dimensions of the wooden plank. On the contrary, the wireless sandals can be used in a wider area, even outdoor, as long as the wireless transmission is ensured (about 20 m from the emitter, although the greater the distance from the emitter the greater the latency). Regarding the microphones-based system, in order to increase the area available for walking more wooden planks and corresponding microphones can be utilized, as long as an identical detection is ensured in all the points of the built floor.

Both the MoCap-based system and the wireless sandals system work independently from the type of floor on which the user walk upon (for instance, there is no difference between walking over floors with and without carpet), and can be used while walking on a treadmill.

On a separate note, a difference exists between the three architectures for what concerns the synthesis of the solid surfaces. While the second and third system make use of the mechanism of triggering prerecorded exciters, the first system uses real footstep sounds as input for the sound synthesis engine. In presence of shoes with a soft sole (e.g., snickers) the simulation of the solid surfaces is less accurate, while using shoes with a hard sole the simulation results more realistic. Nevertheless, the sounds produced by the impact of shoes with hard sole on the wooden plank could become audible in presence of strong steps generating loud sounds.

As far as the audio delivery method is concerned, loudspeakers can be used solely for the MoCap-based system and wireless sandals, while wireless headphones can be involved in all the three apparati.

6.2. Accuracy

The accuracy of the three locomotion interfaces is high in terms of foot–floor interaction types detection, range of the detectable dynamics, and their reproducibility in the sonic simulations. The range of dynamics was measured by recording for different surface materials the sounds generated by stepping strongly and softly, and considering the maximum and minimum peaks in the respective waveforms. During the measurements of the microphones-based system male dress shoes with hard sole were utilized. The same type of shoe was simulated during the recordings conducted using the other two systems. As illustrated in Table 1, the microphones-based system allows to reach the widest range of dynamics in the sonic simulations, on average 3.5 dB larger than the other two systems.

Table 1
Systems range of dynamics (in dB) for different materials. Mic = microphones-based system; MC = MoCap-based system; WS = wireless sandals.

	Mic	MC	WS
Gravel	21	17.5	18
Snow	31	27	27.5
Wood	24	21	22
Metal	21	16.5	17.5
Water puddle	22.5	18.5	19

6.3. Timeliness

All the systems are exempt from latency problems since not noticeable delay occurs between action and sound delivery. The microphones-based system exhibits a latency of about 3 ms due to the auditory feedback synthesis and delivery. As for the MoCap-based system, it is about 6 ms on average (2 ms for the data acquisition and transmission, 1 ms for the real-time data analysis, and 3 ms for the auditory feedback synthesis and delivery). The utilized infrared cameras work at 100 FPS sample rate, which produces 10 ms of latency, as stated by the manufacturer. Such a limit was drastically lowered by triggering the exciter not when the foot actually hit the ground but some ms before (see Fig. 5(a)), i.e., when the foot was still on air and the z-coordinate of the marker became lower than a threshold (around 3 cm from the ground). So the triggering of the exciter was anticipated, and the sound could be delivered to the walker with a negligible latency. On the other hand, the total latency exhibited by the wireless sandals system amounted to about 15 ms at a distance of less than 5 m from the laptop (12 ms for the data transmission and on average 3 ms for the auditory feedback synthesis and delivery).

6.4. Wearability

Both the microphones-based system and the MoCap-based system allow users to wear their own footwear. The only technology required to be worn by the users of the MoCap-based system consists of the four sets of markers which have to be attached to the shoes by means of scotch tape. However, they are very light and therefore their presence is not noticeable, and in addition they do not constitute an obstacle for the user's walk. On the other hand, the microphones-based system requires the use of a wireless headphones set. Conversely, the wireless shoe system is not shoe-independent, and users need to carry the box containing the Arduino board which is attached at the trousers, although its presence is barely noticeable.

6.5. Ecological validity

The involved synthetic auditory stimuli are valid from the ecological point of view as assessed in previous research [21]. Results of an interactive listening experiment showed that most of the surfaces synthesized using the proposed footstep sounds engine was recognized with high accuracy. In particular they were proven to be correctly classified in the corresponding solid and aggregate surface typology. Similar accuracy was noticed in the recognition of real recorded footstep sounds, which was an indication of the success of the proposed algorithms and their control.

As far as the interaction is concerned, different levels of ecology are achieved with the three proposed solutions. Such differences are due to the possibility for the users of wearing their own footwear, the presence of technology attached to the user and the accuracy in the movement dynamics detection and reproduction. The microphones-based system is the most accurate but it requires to wear headphones. Conversely, the other two systems are less accurate but do not require to wear the headphones. However, headphones are recommended in place of loudspeakers especially when the navigation happens in large rooms, in order to achieve a greater sensation of the directionality of the sound as coming from the feet. Therefore, the microphones-based system allows to accomplish the highest ecologically-valid human–system interaction.

6.6. Embodiment

All the developed interfaces allow to achieve an embodied interaction. First of all, the sounds delivered to the users are synchronized with the movements of their feet, resulting into a credible closed-loop interaction. Secondly, no symbolic association are utilized, and as a result the human–computer interaction is intuitive and natural. Thirdly, all the interfaces have a high degree of affordances and allow for direct manipulation.

6.7. Applications

Several are the contexts in which such locomotion interfaces might find application, such as virtual reality, augmented reality, entertainment, training, rehabilitation, or studies investigating the relation between sound perception and action or the influence of the sound on the locomotion. Given the affordance differences, the systems are suitable for different applications and user needs. The microphones-based system is suitable for all the scenarios in which the sonic reproduction of the feet dynamics is the most salient factor. In a different vein, the other two systems are suitable when the user has to be provided with the sonic simulation of a type of shoe different from the one actually worn, as well as when it is important to sonify in a different way the right and left foot, or even the heel and toe. Also, they are suitable in the contexts in which the locomotion has to be analyzed in real-time for providing the user with feedback about the performed movements. In addition, the wireless sandals allow outdoor use, as well as the coexistence of multiple users in the same virtual environment.

All the three systems were developed at software level as extensions to the Max/MSP platform, which can be easily combined with several interfaces and different software packages. As regards the integration in virtual environments, on the one hand the systems allow the simultaneous coexistence of interactively generated footstep sounds and of soundscapes provided by means of wireless headphones (all systems) and of loudspeakers (only the second and third systems). On the other hand, all the architectures can be integrated with visual feedback to simulate different multimodal environments using wireless systems, such as CAVE (Cave automatic virtual environment). Contrary to the microphones-based system and the MoCap-based system, the wireless shoes system could be extended embedding some actuators in the sandals in order to provide the haptic feedback. For this purpose a second wireless device receiving the haptic signals must be involved. Nevertheless, this implementation requires particular care regarding the latency for the round-trip wireless communication.

7. Conclusion

In this paper three custom-made wireless systems able to interactively sonify the foot movements of a walker were presented. In the first system an array of microphones placed on the floor under a wooden plank was utilized. In the second system, the possibilities offered by the motion capture technology were exploited. In the third system, a pair of sandals were enhanced with two force sensitive resistors and two accelerometers for each foot. The main differences among the three systems were based on the wearability, shoe-independence, accuracy in the detection and sonic reproduction of the foot dynamics, and separate tracking of two feet or their parts. Indeed, the microphones-based system was very accurate but it was not easily portable and did not allow the separate tracking of the left and right foot. In addition, it allowed the users to wear their own footwear but better simulations of solid surfaces are achievable only when the user is wearing shoes with hard sole. The MoCap-based system allowed to satisfy the requirement of shoe-independence, allowed the distinct detection of the two feet but was less accurate as far as the foot dynamics are concerned. Moreover, it required a custom made laboratory where to place the cameras. On the other hand, the sandals enhanced with sensors were easily portable, usable also in outdoor environments, but required the users to wear custom made footwear.

The developed locomotion interfaces find application in several contexts, such as augmented reality, virtual reality, or entertainment, as well as in perceptual studies investigating the influence of interactive sounds on locomotion performance usable for training and rehabilitation purposes.

Acknowledgments

The research leading to these results has received funding from the Danish Council for Independent Research Technology and Production Sciences (FTP), Grant No. 12-131985, granted to Luca Turchet. The author gratefully acknowledges Smilen Dimitrov for the help in building the electronics for the wireless sandals, and Prof. Roberto Bresin for providing the equipment to develop the microphones-based system.

References

[1] Rocchesso D, Serafin S. Sonic interaction design. Int J Hum–Comput Stud 2009;67(11):905–6.
[2] Hermann T, Hunt A, Neuhoff J, editors. The sonification handbook. Logos Publishing House; 2011.
[3] Hermann T. Taxonomy and definitions for sonification and auditory display. In: Proceedings of the 14th international conference on auditory display, ICAD; 2008. p. 1–8.
[4] Dourish P. Where the action is: the foundations of embodied interaction. The MIT Press; 2001.
[5] Fontana F, Visell Y, editors. Walking with the senses perceptual techniques for walking in simulated environments. Logos-Verlag; 2012.
[6] Visell Y, Law A, Cooperstock JR. Touch is everywhere: floor surfaces as ambient haptic interfaces. IEEE Trans Haptics 2009;2:148–59.
[7] Turchet L, Serafin S, Dimitrov S, Nordahl R. Physically based sound synthesis and control of footsteps sounds. In: Proceedings of digital audio effects conference; 2010. p. 161–8.
[8] Redd C, Bamberg S. A wireless sensory feedback system for real-time gait modification; 2011.
[9] Redd C, Bamberg S. A wireless sensory feedback device for real-time gait feedback and training. IEEE/ASME Trans Mech 2012;17(3):425–33.
[10] Paradiso J, Hsiao K, Hu E. Interactive music for instrumented dancing shoes. In: Proc of the 1999 international computer music conference; 1999. p. 453–6.
[11] Papetti S, Civolani M, Fontana F. Rhythm'n'shoes: a wearable foot tapping interface with audio-tactile feedback. In: Proceedings of the international conference on new interfaces for musical expression; 2011. p. 473–6.
[12] Morris SJ, Paradiso JA. Shoe-integrated sensor system for wireless gait analysis and real-time feedback. In: Proceedings of the second joint 24th annual conference and the annual fall meeting of the biomedical engineering society engineering in medicine and biology, vol. 3; 2002. p. 2468-9.
[13] Benbasat A, Morris S, Paradiso J. A wireless modular sensor architecture and its application in on-shoe gait analysis. In: Sensors, 2003. Proceedings of IEEE, vol. 2; 2003.
[14] Bamberg S, Benbasat A, Scarborough D, Krebs D, Paradiso J. Gait analysis using a shoe-integrated wireless sensor system. IEEE Trans Inform Technol Biomed 2008;12(4):413–23.
[15] P. Cook, Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds, Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio (2002) 73–78.

[16] Fontana F, Bresin R. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. Proc Colloq Music Inform 2003:109–14.
[17] Farnell A. Marching onwards: procedural synthetic footsteps for video games and animation. In: Proceedings of the pure data convention.
[18] Fontana F, Morreale F, Regia-Corte T, Lécuyer A, Marchal M. Auditory recognition of floor surfaces by temporal and spectral cues of walking. In: 17th International conference on auditory display; 2011.
[19] Turchet L. Footstep sounds synthesis: modelling foot–floor interactions, shoes, grounds, and the walker. In: ACM transactions on applied perception; submitted for publication.
[20] Turchet L, Nordahl R, Berrezag A, Dimitrov S, Hayward V, Serafin S. Audio-haptic physically based simulation of walking on different grounds. In: Proceedings of IEEE international workshop on multimedia signal processing. IEEE Press; 2010. p. 269–73.
[21] Nordahl R, Serafin S, Turchet L. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proceedings of the IEEE virtual reality conference. IEEE Press; 2010. p. 147–53.
[22] Serafin S, Turchet L, Nordahl R, Dimitrov S, Berrezag A, Hayward V. Identification of virtual grounds using virtual reality haptic shoes and sound synthesis. In: Proceedings of Eurohaptics symposium on haptic and audio–visual stimuli: enhancing experiences and interaction; 2010. p. 61–70.
[23] Turchet L, Serafin S, Nordahl R. Examining the role of context in the recognition of walking sounds. In: Proceedings of sound and music computing conference; 2010.
[24] Turchet L, Serafin S, Dimitrov S, Nordahl R. Conflicting audio-haptic feedback in physically based simulation of walking sounds. In: Haptic and audio interaction design. Lecture notes in computer science, vol. 63. Berlin, Heidelberg: Springer; 2010. p. 97–106.
[25] Gibson J. The senses considered as perceptual systems. Houghton Miffin; 1966.
[26] Gaver W. What in the world do we hear?: an ecological approach to auditory event perception. Ecol Psychol 1993;5(1):1–29.
[27] Gaver W. How do we hear in the world? Explorations in ecological acoustics. Ecol Psychol 1993;5(4):285–313.
[28] Serafin S, Turchet L, Nordahl R. Extraction of ground reaction forces for real-time synthesis of walking sounds. In: Proceedings of the audio mostly conference; 2009. p. 99–105.
[29] Peltola L, Erkut C, Cook P, Valimaki V. Synthesis of hand clapping sounds. IEEE Trans Audio, Speech, Lang Process 2007;15(3):1021–9.
[30] Avanzini F, Rocchesso D. Modeling collision sounds: non-linear contact force. In: Proceedings of digital audio effects conference; 2001. p. 61–6.
[31] Avanzini F, Rath M, Rocchesso D. Physically-based audio rendering of contact. Proceedings IEEE international conference on multimedia and expo. ICME'02, vol. 2. Berlin: IEEE; 2002. p. 445–8.
[32] Avanzini F, Serafin S, Rocchesso D. Interactive simulation of rigid body interaction with friction-induced sound generation. IEEE Trans Speech Audio Process 2005;13(5):1073–81.
[33] Bresin R, Delle Monache S, Fontana F, Papetti S, Polotti P, Visell Y. Auditory feedback through continuous control of crumpling sound synthesis. In: Proc CHI workshop on sonic interaction design; 2008. p. 23–8.
[34] Lyon E, Penrose C. Fftease: a collection of spectral signal processors for max/msp. In: Proceedings of the international computer music conference. San Francisco: International Computer Music Association; 2000. p. 496–8.