Applied Acoustics 107 (2016) 46–68

Contents lists available at ScienceDirect

Applied Acoustics

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

Footstep sounds synthesis: Design, implementation, and evaluation of foot–floor interactions, surface materials, shoe types, and walkers' features

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 26 November 2014
Received in revised form 13 May 2015
Accepted 22 May 2015
Available online 17 July 2015

Keywords:
Footstep sounds
Physical models
Sound synthesis

ABSTRACT

This paper presents a novel footstep sound synthesizer based on physical and physically inspired models coupled with additive synthesis and signals multiplication. Several types of foot–floor interactions are simulated (e.g., different types of steps in walking and running or the sliding of the foot on the floor). Moreover, different types of shoes and ground materials (solid, aggregate, liquid, and hybrids) are synthesized, along with the modeling of some anthropomorphic features of the walkers (i.e., body size and foot-length). The design choices underlying the proposed synthesis methods were made according to four main points: (i) auditory perceptual relevance, i.e., ecological validity; (ii) cartoonification approach; (iii) parametric temporal control; (iv) real-time utilization. Moreover, four types of control for the involved synthesis algorithms are discussed. Firstly, a control strategy is proposed in order to generate sequences of footstep sounds. Secondly, the design choices underlying the tuning of the synthesis parameters are illustrated. Thirdly, a control strategy is presented to provide footstep sounds designers and foley artists with a tool to create perceptually compelling sounds in an intuitive manner. Fourthly, control techniques are discussed for the interactive case in presence of different types of locomotion interfaces along with their differences with the non-interactive control when locomotion is passively simulated. Finally, four perceptual experiments successfully assessed the validity of the proposed techniques.

1. Introduction

Footstep sounds represent important elements in the movie and computer games industry [1]. Usually such sounds are acquired from sound libraries or recorded by the so-called "foley artists" who simulate the locomotion of the actor using different techniques, which can even avoid the use of shoes and surface materials [2].

A footstep sound is the result of the interaction between the foot and the floor. It depends on the type of shoe, kind of surface material, dynamics and temporal evolution of the contact between the shoes and the ground, and person's anthropomorphic features (such as height, weight and foot-length). In particular, ground materials may be classified in four typologies: solid, i.e., hard and homogeneous materials (e.g., asphalt, wood); aggregate, i.e., materials possessing a granular structure (e.g., gravel, snow); liquid, i.e., viscous materials (e.g., water puddles); hybrid, i.e., materials encompassing characteristics of more than one of the previous typologies (e.g., mud, wet gravel, wet concrete).

Recently the interest for simulating footstep sounds algorithmically has grown, especially in virtual reality contexts [3]. The first systematic attempt to synthesize the sounds of people walking on different aggregate surfaces was proposed by Cook [4]. That research originated in 1997 from a collection of physically informed stochastic models (PhISM) used to simulate several musical and everyday sonic events [5]. A similar approach was also proposed by Fontana and Bresin in 2003, where physically-based algorithms for the reproduction of microscopic impacts were driven by a stochastic controller in order to simulate footsteps over crumples and similar aggregate materials [6]. In addition, the continuous control of physical floor parameters and users' gestural intentions was achieved by means of a preliminary model for expressive control and grouping of the synthesized sounds into footstep sequences. Based on that work, De Witt and Bresin proposed a preliminary model where the footsteps synthesis was malleable to the emotional influence of the user manipulating the interface [7].

Procedural sound synthesis of walking was also proposed by Farnell, where the characteristic events of a footstep sound were reproduced by simulating some biomechanical parameters of

E-mail address: tur@create.aau.dk

http://dx.doi.org/10.1016/j.apacoust.2015.05.013
0003-682X/

locomotion [8]. Precisely, a synthesis algorithm was controlled by means of a mathematical simulation of the so-called ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step [9]. This approach allowed for real-time modification of the walkers' speed and weight, as well as the material composition of the ground.

In 2010, a footstep sound synthesizer based on physical models was proposed by Turchet et al. to simulate the sounds of walking on solid and aggregate surfaces, along with locomotion interfaces for its real-time control [10]. In 2011, by exploiting some of the techniques presented by Cook [4], Fontana et al. proposed a synthesis method based on a LPC analysis-resynthesis technique [11].

The main goal of the present research was to provide footstep sounds designers and foley artists with a tool to create, in an intuitive manner, perceptually compelling sounds valid from the ecological point of view [12–15]. Until now research on footstep sounds synthesis has focused only on the rendering of the ground material. No attention has been devoted to the simulation of the shoe type, anthropomorphic features of the person performing the locomotion, or foot–floor interactions (such as, for instance, the sliding of the foot on the floor or the exact dynamics with which the foot hits the ground). It is the author's conviction that the synthesis of a sound resulting from the interaction of the feet with the floor cannot avoid considering simultaneously all the factors which contribute to it. Therefore, a holistic approach to the footstep sound synthesis was followed. A physically-based sound synthesis engine, which extends that described in [10], is proposed to simulate (i) a larger palette of ground materials; (ii) the shoe type; (iii) the type of interaction of the foot with the floor; (iv) the person's anthropomorphic features; (v) the person's locomotion. To develop the synthesis engine tool, as well as the models underlying it, the analysis-by-synthesis method was applied [16,17]. Such a method consists in designing a model by gathering knowledge about it from data collected via measurements, interviews to experts, or hypotheses, and by successively verify its validity through synthesis (i.e. by implementing the model in a software tool) and psychophysical tests.

The developed synthesizer has been conceived to be used interactively in conjunction with a locomotion interface [10,18], as well as in non-interactive contexts where the locomotion is simulated at auditory level without involving physical actions.

2. Design choices

2.1. The ground truth problem

The first issue encountered in the synthesis process was the lack of a comprehensive theoretical framework of auditory recognition of footstep sounds in all their possible combinations of materials, shoes, types of locomotion, and person's anthropomorphic features. Indeed, only a handful of studies investigated the auditory recognition of locomotion sounds, and almost all in walking contexts. Therefore, a complete "ground truth" on the recognition of natural or synthetic footstep sounds that could be used as a guideline for the design of the synthesis algorithms, as well as a reference point for evaluating their quality, was missing. Hereinafter, the results of perceptual experiments conducted in both natural and synthetic settings are briefly summarized.

Following the principles of ecological psychology applied to sound and hearing [12,13], research on perception of locomotion sounds has studied different properties of walking sound sources. Li et al. investigated the ability of subjects to identify the gender of a walker by listening to his/her footstep sounds on hardwood and using leather sole shoes [19]. Results showed that the gender of most of the walkers was identified at better than chance level. Also, results showed that maleness judgments were related to the anthropomorphic features of the walkers such as height, weight and foot-length, as well as they were based upon spectral differences between male and female walkers. Specifically, the spectrum of males' footsteps was characterized by a spectral mode (i.e., the most prominent spectral frequency) that was placed at lower frequencies and was narrower compared to the females' one. Also, the contribution of the high frequencies was different in the two genders: the spectrum of males' footstep sounds was more skewed toward low frequencies and had a more rapid fall. Nevertheless, the major role in the classification of the walker gender was played by the spectral peak. This is consistent with the fact that the fundamental frequency is usually determined by the size of the source [20]. Furthermore, it was found that a faster walk was more likely to be judged as a female.

The gender recognition was studied also in the context of synthetic footstep sounds. In the interactive listening tests reported by Visell et al. [21], which made use of the footstep synthesizer proposed by De Witt and Bresin [7], subjects could adjust pace and material to determine the gender of a virtual walker. Results showed that subjects associated both different pace and material to the two genders: female walkers were identified by faster pace and by materials with higher resonant frequency. Such results are perfectly in agreement with those reported by Li et al. [19]. A subsequent experiment showed also the ability of subjects of interactively creating walks with emotional intentions by varying temporal and sound level parameters. Such an experiment was designed on the basis of the results presented by Giordano and Bresin [22] that studied the recognition of emotions from the auditory information contained in sequences of real footstep sounds corresponding to walks performed with different emotional intentions. Those results showed the presence of strong similarities between walking and musical expression of emotions with respect to acoustical variables such as temporal evolution and sound level. These findings were furthermore confirmed by two studies involving instrumented shoes to generate interactively synthetic footstep sounds [23,24]. Moreover, the gender identification was associated with the footstep sounds' spectral content consistently with the findings reported in [19].

In a different vein, Pastore et al. showed that listeners are capable of identifying the posture of the walker who generated the acoustic stimuli [25], while Mäkelä et al. showed that footstep sounds can convey information about the identity of a person [26].

On a separate note, Giordano and colleagues studied the recognition of walked-upon ground materials [27]. Participants were presented with recordings of their own walking sounds using rubber-sole shoes on solid (vinyl, wood, ceramic, marble) and aggregate materials (very small gravel, small gravel, medium gravel, and large gravel), and were asked to identify them in a forced choice task with the same eight materials. Results showed that the material identification was not very accurate, but that aggregate materials were seldom identified as solids and vice versa.

Along the same line, Nordahl et al. investigated the ability of listeners of identifying synthetic and real walked-upon surface materials [28]. Three groups of subjects were involved. The first group passively listened to recordings of real footsteps. The second group interactively generated synthetic footstep sounds simulated by means of the synthesis engine described by [10]. The third group passively listened to pre-recorded footstep sounds produced by the same synthesis engine. Results showed that subjects were able to recognize most of the synthesized surfaces with high accuracy. In particular, the proposed simulations were proven to be correctly classified in the corresponding solid and aggregate surface typology. Similar accuracy was noticed in the recognition of the recordings of real footstep sounds, which was an indication of the success of the proposed algorithms and their control. More importantly, the study provided interesting insights in how multimodal interaction affects auditory material perception: in the interactive setup, subjects were able to identify synthesized surface materials at a comparable accuracy with real-world recordings, while the performance with pre-recorded sounds was significantly worse than the other two.

The auditory recognition of two forms of locomotions, namely walking and running, was studied in [29]. The stimuli were sequences of recorded footsteps sounds of a man walking or running on gravel, single footstep sounds extracted from those sequences, sequences of footsteps obtained by looping the same footstep sound, and sequences of footsteps obtained applying some of the musical performance rules reported by Friberg [30]. Subjects were asked to identify whether the stimuli corresponded to a walking or running sound, and whether the sound was human or mechanical. Results showed better than chance identification of the two form of locomotion for all stimuli. However, stimuli corresponding to a single footstep were classified with less precision. In addition, the sequences of footstep sounds produced by the control models were all classified as mechanical.

Finally, the study reported by Turchet and Serafin [31] investigated the role of temporal and amplitude aspects in sonically simulating the act of walking on bumps, holes and flat surfaces. In particular, it was investigated whether the timing between heel and toe and the timing between footsteps, as well as variations in sound level of each step in the sequence, affected perception of walking on uneven surfaces. Results showed that it is possible to sonically simulate those three types of surface profiles only by varying temporal information.

To the author's best knowledge, to date, there are several aspects of locomotion sounds that have never been studied. Examples are the identification of sounds resulting by walking or running with different types of shoes on various materials, by jumping actions, or by different types of foot–floor interactions. However, the construction of a comprehensive theoretical framework of auditory recognition of real footstep sounds that could serve as a ground truth to compare the quality of the corresponding synthesized sounds poses several problems. First of all, to date a comprehensive database of recordings of real footstep sounds usable for this kind of study is missing. Indeed, despite it is possible to find several repositories of footstep sounds, the information available for each sound excerpt is not complete. Most of the times the information regarding the surface material is present but not that about the type of shoes, or that about the gender of the walker is available but not his/her anthropomorphic features. More importantly, insufficient information is provided regarding whether the sounds have been recorded from real footsteps or from footsteps created by means of foley artists' techniques. Moreover, no information whatsoever is specified about the temporal evolution and dynamics of the foot–floor contact for each step. In addition, the information concerning the dimensions and properties of indoor environments which can lead to reverberations that can even affect the spectral content of a sound is almost always unknown. Furthermore, the available sounds have been recorded using different microphones, and the majority of the times no information is available about the distance of the microphone from the sound source, or about the actual sound level of each step.

To build a ground truth for footstep sounds one should study, in a controlled way, all the factors contributing to the sound as well as their mutual interaction (e.g., to study the effect of different types of shoes on the auditory perception of a same ground material). However, this is hardly possible not only for all the reasons described above, but also for the following ones. Recordings of real footstep sounds almost always also contain auditory information concerning the environment (e.g., birds chirping, wind, rain falling) which could bias the perception of the factors contributing to the footstep sounds (e.g., the ground material), as demonstrated by Turchet et al. [32]. So, one would need recordings of footstep sounds devoid of contextual information. This might be possible by extracting the sound of a single footstep as pure as possible, and concatenating it to form a walking or running sequence. However, this approach is not valid from the ecological point of view and listeners judge those sequences as mechanical, as demonstrated in [29]. Therefore, this could affect the quality of the results of a listening test investigating the various aspects of footstep sounds (e.g., realism). On the other hand, an investigation conducted on isolated footsteps is also lacking of ecological validity and might lead to unnatural and even confusing perceptions (e.g., the sound of a single isolated female footstep on wood using high heels could be identified as a book being dropped on a table). However, using recorded sequences of footstep sounds poses the problem that the speed of the walkers/runners/jumpers should be kept constant in order to study the effect of other factors contributing to the sound. Nevertheless, the available recordings greatly differs in terms of locomotion speed and such differences might influence the perception of some characteristics of the source generating the footstep sounds, such the gender, as demonstrated in [21]. Finally, the common everyday experience suggests that reverberation plays a relevant role in the perception of surface materials: a same solid material impacted in exactly the same way by a walker could be perceived differently depending on the size and shape of the reverberating room, since these are factors that might greatly alter the temporal and spectral content of the sound. Similarly, the accuracy in the source identification may depend on the distance of the listener [33].

2.2. The plausibility objective

To cope with the ground truth problem, the approach to the synthesis adopted here considered the footstep sounds' general properties that are relevant to auditory perception. According to the ecological acoustics [12–15], the direct perception of the features of a sound source is based on the so-called "invariants", i.e., properties of the acoustical signal that specify a given feature of an event despite variations in other features of the sound source. The invariants can be grouped into two categories: (i) structural invariants, which specify the intrinsic properties of an object (e.g., size, shape, material) and enable us to recognize it; (ii) transformational invariants, which characterize external interactions on an object (e.g., impacts, frictions) and enable us to recognize the actions that produced the sound.

However, to recognize a sound, listeners do not rely only on such pieces of acoustic information that invariably specify the properties of the object and the action that caused that sound. Indeed, as demonstrated by the findings of a series of experiments reported by Ballas [34], the listeners' identification performance is influenced by other factors besides the acoustical variables, such as ecological frequency (i.e., the frequency with which a listener encounters a specific sound event in everyday life) or causal uncertainty (i.e., the amount of reported alternative causes for a sound). In more detail, acoustical variables accounted for only about half of the variance in identification accuracy.

Along the same line, Lemaitre et al. identified three properties reported by listeners when describing a sound: acoustic properties, causal properties (i.e., invariants of the sound source), and semantic properties (related to the interpretation of the source) [35]. Their study revealed that the information that listeners focus on depends on both the listener's expertise and the identifiability of the sound. Interestingly, it was found that non-expert listeners tend to focus more spontaneously on the causal properties of the sound event, i.e., on its structural and transformational invariants.

Another aspect that can influence the identification of a sound source is the context. For instance, this has been proved for the case of footstep sounds. The study reported by Turchet et al. [32] investigated the role of contextual information, sonically provided as soundscape, on the perception of synthesized footstep sounds. Subjects were asked to walk inside a limited area in a laboratory, while simulations of footstep sounds on different surfaces were interactively generated with and without an accompanying soundscape. Soundscapes sonically simulated either the environment typically associated with the surface material synthesized (i.e., coherently), or with an atypical one (i.e., incoherently). Results showed that, in some conditions, adding a coherent soundscape significantly improved both recognition of surface materials and realism evaluations when compared to both footstep sounds alone and with an accompanying incoherent soundscape.

The objective of the present research was to create algorithms capable of simulating footstep sounds in a perceptually compelling way, having as a reference the work of foley artists who invent, with their creativity, methods to produce sounds that appear plausible for a certain scene displayed on screen, i.e., semantically coherent with a contextual information and in line with general expectations [2]. To achieve this goal in absence of a comprehensive ground truth that could be used as a guideline for the design of the synthesis algorithms, an invariants-based design strategy was adopted. It consisted of a two phases process: (i) the analysis of the recordings of real footstep sounds, both at acoustic and perceptual level, aiming at finding their perceptual invariants; (ii) a synthesis of such invariants. On the basis of the structural and transformational invariants, a synthesis paradigm was proposed in which the sound was defined as the result of an interaction between the shoe and the floor. In this paradigm, the floor's properties were separated from the interactions it was subjected to, while the shoe's properties were in part separated and in part linked to the interaction.

The goal of the analysis phase was to individuate both structural and transformational invariants characterizing footstep sounds for each combination of materials, shoe types, foot–floor interactions, and person's anthropomorphic features. Particular attention was given to those features simply recognizable to a first listening. For example, the acoustical signal corresponding to a step generated by a male walking in shoes with hard soles and squeaking material upon a concrete floor is characterized by two subcomponents, i.e., the impact between two solid objects having a certain level of hardness, and the friction of the shoe material. In addition, the heel and toe contributions can be distinguished, as well as their different forces of impact. The impact between these kinds of shoes and floor results in a sound with faster attack and higher amplitude than that achievable using a rubber sole shoe when exerting the same impact force.

In the synthesis phase, the structural invariants of both shoes and floor and the transformational invariants of the foot–floor interactions were considered as subcomponents characterizing the sound itself, and were then simulated independently by means of algorithms capable of modeling them. Subsequently, they were combined appropriately in order to construct the wanted global sound. In particular, the amplitude of the different subcomponents was weighted according to a similar contribution present in the corresponding real sound to be simulated. However, the overall amplitude of each simulation was not chosen in order to reproduce that of real life sounds. Indeed, the paradigm of movies and computer games was followed, where sounds are usually enhanced in amplitude in order to capture the user's attention [36]. In this regard, results of a previous experiment about amplitude perception of synthesized footstep sounds were adopted as a guideline for the amplitude settings [37].

On the one hand, such an approach to the synthesis based on the separate rendering of the various subcomponents in a footstep sound, was inspired by the organizing idea in auditory display research of decomposing complex everyday sound phenomena in terms of more elementary ones [12,38]. On the other hand, it was inspired by research on sound source perception that showed that perceptual judgments integrate information from multiple acoustical features [39]. To this regard it is important to highlight that the proposed approach is robust enough from the point of view of the sound delivery: loudspeakers and headphones inevitably alter the frequency content of a sound while reproducing it.

Moreover, the synthesis considered not only all the perceptual results reviewed above, but also other results obtained in different domains of auditory perception (e.g., impacts between real or synthetic objects [40]). Furthermore, the physical processes occurring during objects collisions were taken into account, as well as several indications suggested by the common everyday experience.

However, the synthesis did not aim to reproduce the exact physical processes occurring for any sort of foot–floor interaction, material, etc. The cartoonification approach was instead followed, i.e., the simplification of the underlying physics and emphasis on the main acoustic features, able to express ecological attributes of the simulated sound source [41]. Such an approach not only allows to simplify the sound models while retaining perceptual invariants, but also to achieve high computational efficiency, which is aimed for interactive contexts such as that of physically navigating in a virtual environment.

In summary, the design choices underlying the synthesis methods proposed here were made according to four main points: (i) auditory perceptual relevance, i.e., ecological validity; (ii) cartoonification approach; (iii) parametric temporal control ensuring appropriate articulations of subcomponents present in the footstep sound; (iv) real-time utilization.

2.3. Footsteps sound analysis

Hundreds of footstep sounds were collected from different free repositories and commercial sound libraries[1] and were analyzed in order to determine the transformational invariants of foot–floor interactions, and the structural invariants of surface materials, shoe types, and person's anthropomorphic features.

As far as the foot–floor interactions are concerned, two invariants were individuated (i.e., those properties characterizing the footstep sound independently from the type of shoe utilized, from the type of floor material, and from the anthropomorphic features of the person): the impact and the friction of the shoe on the floor. During the locomotion (in its three most common forms, walking, running and jumping), footstep sounds are produced as the result of shoe-floor impacts when the swing foot is planted. In addition, depending on the walker's gait, other slithering sounds can be generated when the shoe brushes more or less strongly against the floor before the normal footfall ("scuffs" in foley artists' jargon). Moreover, sounds can be created even when the locomotion is not happening, like for example when cleaning the shoes on a carpet in front of a door or slipping on a downhill surface. The synthesis of such invariants is described in Section 3.1.

As for the structural invariants of the ground materials (i.e., those properties characterizing the footstep sound independently from the type of shoe utilized, from the type of foot–floor interaction, and from the anthropomorphic features of the person), they were identified according to the material typology. For solid materials: stiffness (i.e., the level of hardness of a solid floor) and creakiness (e.g., the creaking of a parquet); for aggregate materials: homogeneity (i.e., the material is composed of the same substance or not), compliance (i.e., to what extent the foot can sink into that

[1] E.g., http://www.freesound.org/, http://www.sounddogs.com/, http://www.hollywoodedge.com/.

material), granularity (e.g., colliding pebbles of gravel), and fracture (e.g., branches breaking); for liquid materials: viscosity, density, and depth of the liquid; for hybrid materials: the combination of the previous invariants. The synthesis of such invariants is described in Section 3.2.

The structural invariants of the shoes materials (i.e., those properties characterizing the footstep sound independently from the type of surface material, from the type of foot–floor interaction, and from the anthropomorphic features of the person) were: the sole hardness, the squeakiness of the shoe material (i.e., the ability of emitting a squeaking sound), its clickiness (i.e., the ability of emitting a clicking sound), the presence of a metallic component (typically a buckle or the spurs of cowboy boots). The synthesis of such invariants is described in Section 3.3.

Finally, the main structural invariants of the body of the person performing the locomotion (i.e., those properties characterizing the footstep sound independently from the type of surface material, from the type of foot–floor interaction, and from the type of shoe) were identified as height, weight and foot length, according to the findings reported by Li et al. [19]. The synthesis of such invariants is described in Section 3.4.

3. Sound models

3.1. Modeling the foot–floor interactions

The feet can interact with the ground in several ways, producing, as a consequence, a large variety of acoustic events of different nature [21]. Therefore, when aiming at designing algorithms to simulate all these interaction possibilities, a general method is desirable.

In the proposed simulations a footstep sound was considered as the result of the interaction between an "exciter", which modeled the contribution to the sound given by the shoe (i.e., type of shoe and type of foot–floor interaction), and a "resonator", which modeled the ground contribution (i.e., the surface material). In particular, the exciter consisted of a signal in the audio domain, while the resonator consisted of several physical models.

In previous research [10,18], the exciter corresponded to the amplitude envelope extracted from an audio signal containing a footstep sound. Such a signal was provided in real-time according to the utilized locomotion interface (typically microphones detecting the walker's footstep sounds or shoes enhanced with sensors triggering pre-recorded audio files containing sounds of footsteps). The envelope (e) was extracted from the signal (x) by means of the non-linear low-pass filter proposed in [42] and subsequently utilized in [4]:

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

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 extracted from recorded footstep sounds on a concrete floor.

Conversely, in the current technique the exciter signal is built from scratch in order to approximate the envelope extracted from real footstep sounds corresponding to different types of foot–floor interaction (see Fig. 2). The core idea is that such interactions and, partially, shoe types (see Section 3.3) can be uniquely described

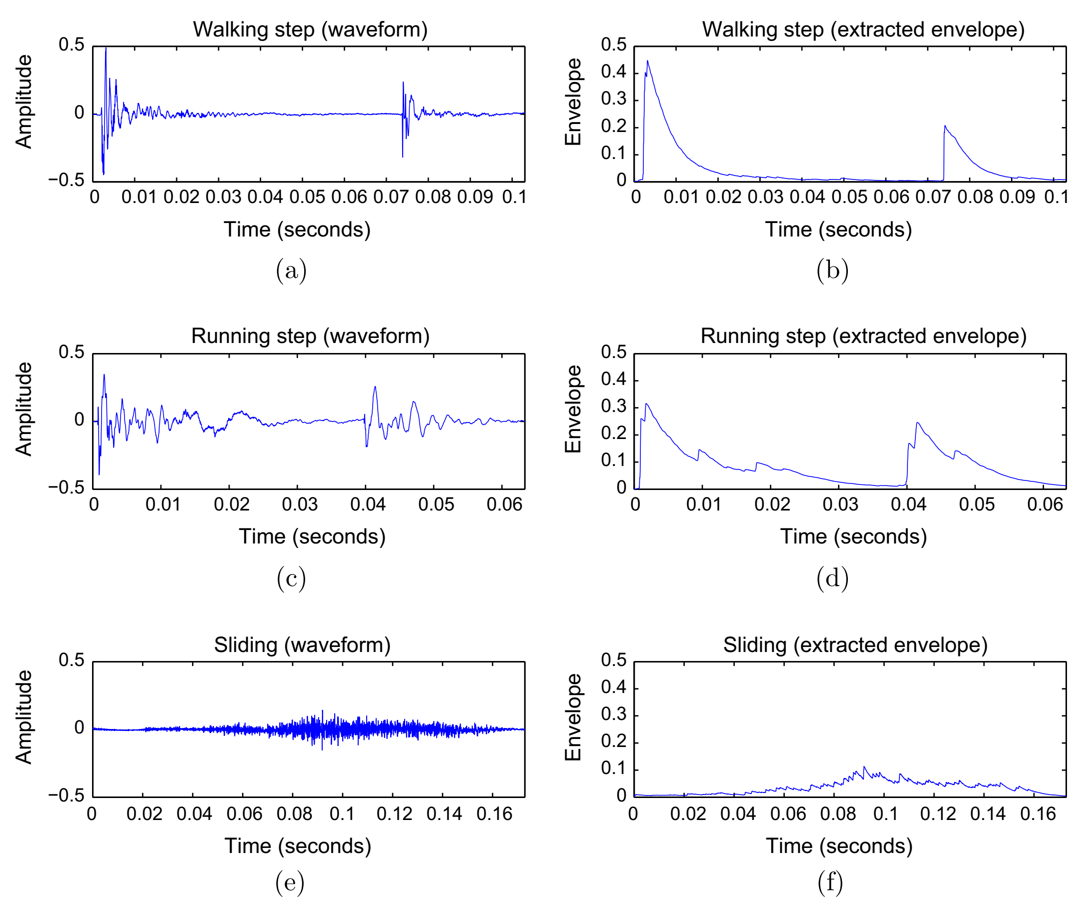

Fig. 1. Waveforms and corresponding amplitude envelopes of typical footstep sounds produced on a concrete floor with male dress shoes during various foot–floor interactions.

and rendered by such a signal by controlling its temporal evolution (e.g., type of attack, decay, peak shape). Such an exciter is also considered to express the vertical component of the GRF.

The two types of identified transformational invariants, impacts and frictions, generate very different acoustical signal, and consequently can be modeled with different types of exciters. For instance, a footstep sound can include the contribution of the heel and of the toe (see Fig. 2(a), (c), and (d)), or can be characterized by only one impulsive signal which comprises both the case where heel and toe simultaneously strike the floor and the case where one of the two does not produce an audible sound (see Fig. 2(b)). Jumping and running steps are characterized by shorter durations and higher amplitudes than those of the steps occurring during walking (see Fig. 2(e)–(h)), as demonstrated by studies on measurements of durations and amplitudes of the GRF in these three types of locomotion [43–45]. Jumping steps differ from running steps manly for the basically simultaneous impact of heel and toe on the floor and for a greater amplitude.

On a separate note, a footstep sound can encompass sub-events due to the brushing of the heel against the floor before impacting it

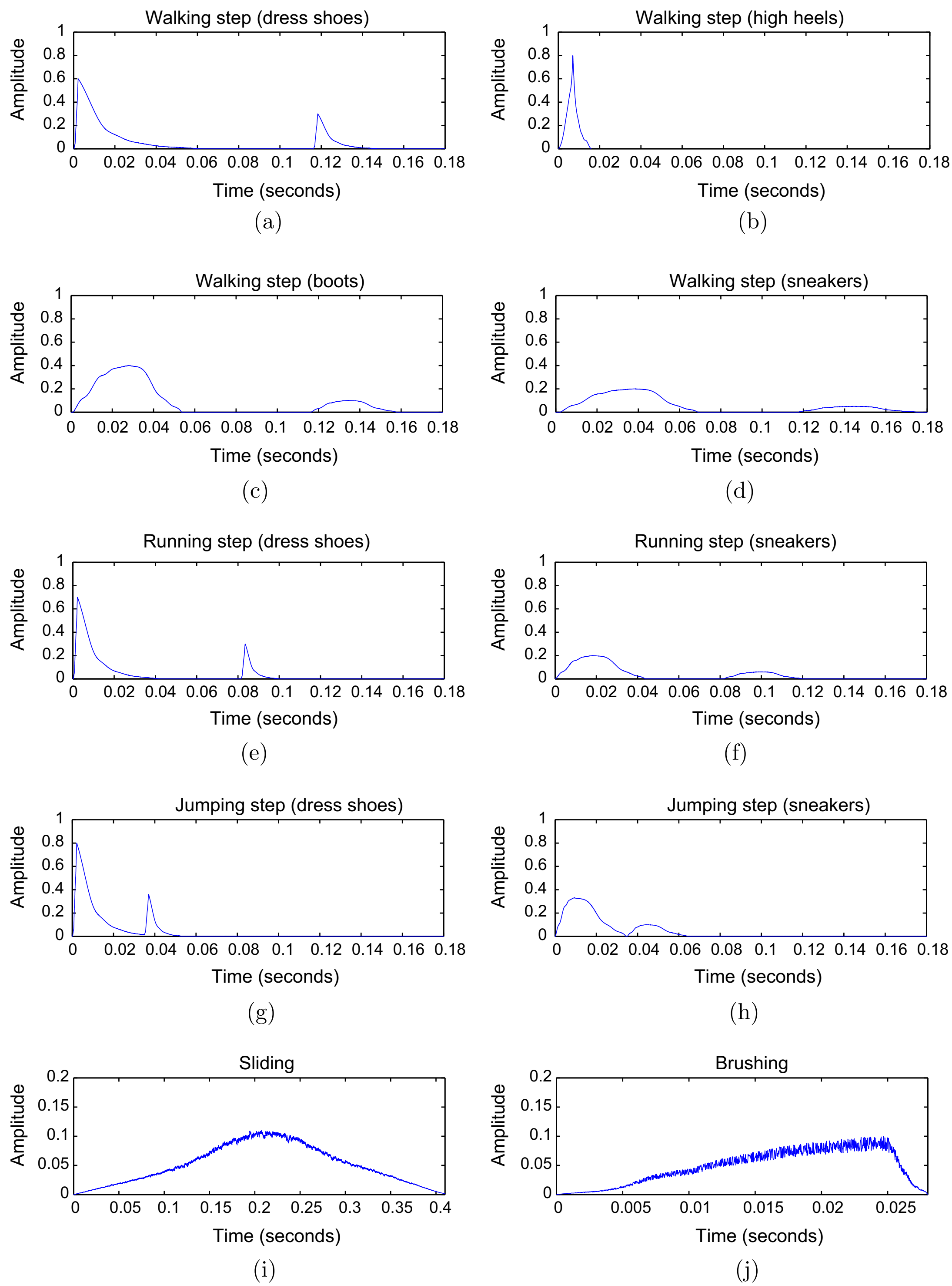

Fig. 2. Exciters corresponding to various foot–floor interactions using different shoe types.

(see Fig. 2(j)), or to the slipping of the toe after its strike. A sliding interaction presents similar characteristics but lasts much longer (see Fig. 2(i)).

3.2. Modeling the ground contribution

To synthesize solid, aggregate, liquid and hybrid surface materials several sound models were utilized. Such models are briefly summarized below, with particular regard to their parameters used to control the synthesis in the proposed simulations. The choice of which sound model to use for each surface was driven by its capabilities of rendering the structural invariants individuated in the analysis phase (see Section 2.3).

Fig. 3 shows the waveforms of the synthesized footstep sounds corresponding to materials belonging to the three different surface typologies.

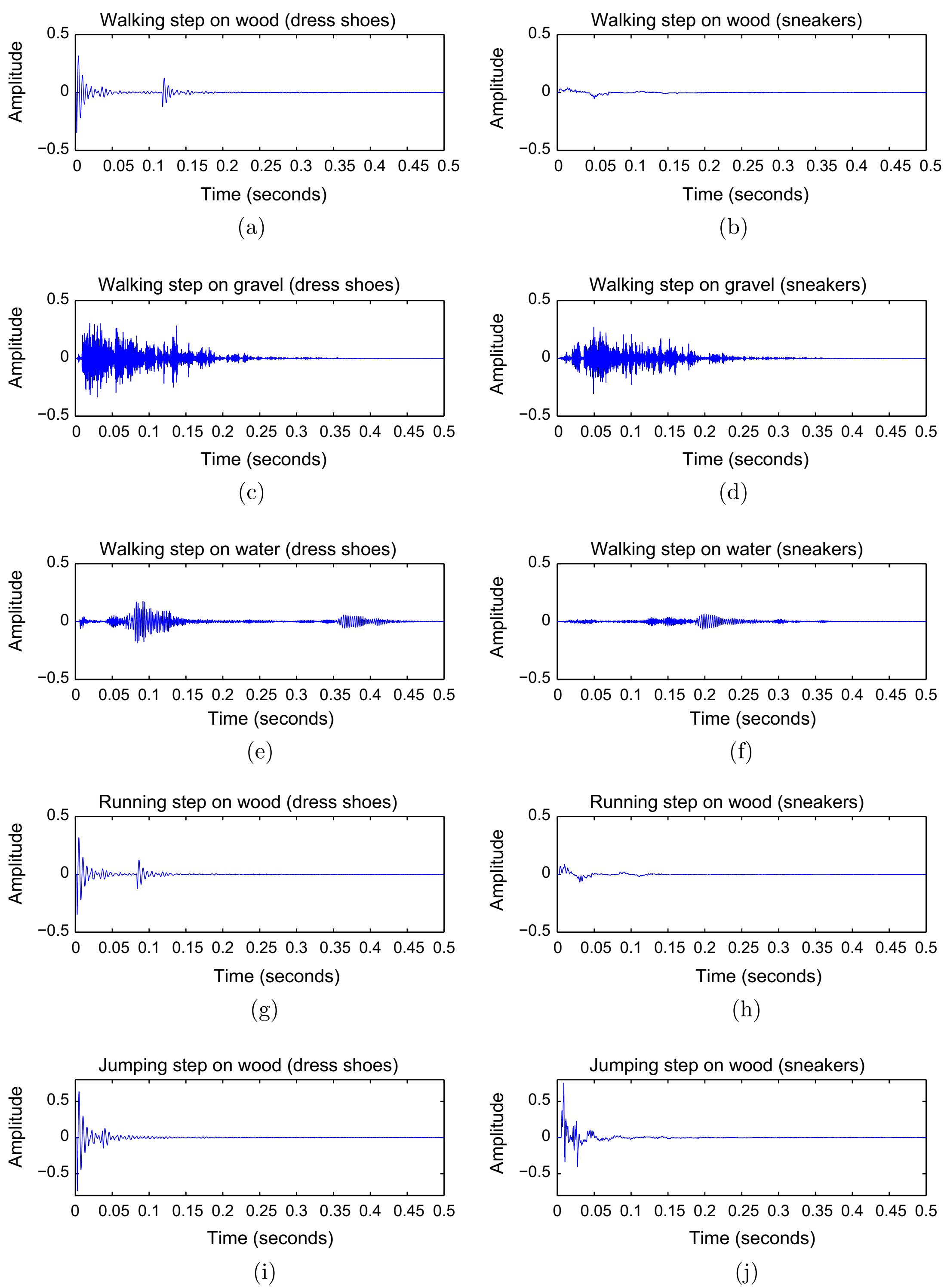

Fig. 3. Waveforms of the synthesized footstep sounds resulting from various combinations of foot–floor interactions, surface materials, and shoe type. The sounds are generated using the corresponding exciters illustrated in Fig. 2.

3.2.1. Physically informed stochastic model

At the heart of the PhISM algorithm [5] are particle models, that is, models characterized by basic Newtonian equations governing the motion of point masses which when colliding produce a sound. The PhISM algorithm simulates particle interactions by using a stochastic parametrization thereby avoiding modeling the collision of each of many particles explicitly. Instead, the particles are assigned a probability to create an acoustic waveform. In the case of many particles, the interaction can be represented using a simple Poisson distribution, where the probability of sound production is constant at each time step, giving rise to an exponential probability weighting time between events.

The model utilizes a second order resonator (biquad filter) which is excited by a signal expressing the input system energy. Other parameters involved in the synthesis control are the resonator frequency, the pole radius, the number of colliding particles, the sound decay of each collision, and the total system decay.

3.2.2. Impact model

The impact model [46,47] simulates the collision between two modal resonators, i.e., objects described as a system of a finite number of parallel mass-spring-damper structures [48]. Each of these structures models a damped mechanical oscillator, which represents a normal mode of resonance of the object. The oscillation period, mass and damping coefficient of each oscillator correspond respectively to the frequency of resonance, gain and decay time of each mode.

The utilized implementation of the impact model [49] has two types of control strategies: a signal expressing velocity and a signal expressing force. In both cases the model's output represents the vibrations of the excited object. When the model is controlled "in velocity" the contact force f between the two objects is modeled by a Hunt-Crossley-type interaction [50], which includes both an elastic component and a dissipative term:

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

In this equation, x represents the contact interpenetration (when $x > 0$ the two objects are in contact), $\dot{x}$ is the compression velocity, k accounts for the material stiffness, λ represents the force dissipation due to internal friction during the impact, and α is a coefficient which depends on the local geometry around the contact surface. For the purposes of the simulations reported here, the exciter object was modeled as an "inertial hammer" (i.e., an inertial mass described with one mode, zero spring constant and zero internal damping), while the excited object as a modal resonator having six modes.

When the model is controlled "in force" the above equation reduces to $f = f_{ex}$, where f_{ex} is the force exerted on the modal resonator that uniquely defines a combination of α, λ and k. Therefore, this force substitutes to the hammer in its function of being the exciter for modes of the resonator.

Furthermore, a white noise burst was added to the sound in correspondence of the attack of both heel and toe strikes since the analysis of footstep sounds on solid surfaces revealed the presence of a noisy component in those two parts. Specifically, the amplitude of such a noisy component was proportional to the amplitude of the exciter.

3.2.3. Friction and fractal noise models

The friction model [51] models the non-linear interaction force that arises during friction between two modal resonators. The relationship between relative velocity v of the bodies in contact and friction force f is represented through a differential equation. Assuming that friction results from a large number of microscopic elastic bonds, called "bristles" [52], the v-to-f relationship is expressed as:

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

where z and $\dot{z}$ are the average bristle displacement and velocity, the coefficient σ_0 is the bristle stiffness, σ_1 the bristle damping, and the term $\sigma_2 v$ accounts for linear viscous friction. The fourth component $\sigma_3 w$ relates to surface roughness, and is simulated by means of a fractal noise model. Roughly, the fractal noise is obtained by filtering white noise in order to generate a noisy sound with spectrum equal to $1/f^{\beta}$, where β is a parameter (called "fractal dimension") related to the "roughness" of the fractal process. Such spectrum can be approximated by a cascade of first-order filters, and varying β it is possible to render textures of surfaces with variable degree of roughness (for more details see [53]).

In the proposed simulations, the first object (the rubber) was modeled as an inertial point mass used as exciter for the second object (the resonator). The only parameter describing the rubber was its mass, while the second object was rendered using two modes. Moreover, the friction model is parametrized by several other quantities, such as the dynamic-friction coefficient, the static-friction coefficient, the break-away coefficient, the Stribeck velocity, the perpendicular pressure that the rubber applies on the resonator, and the external force on the rubber.

3.2.4. Crumpling model

The crumpling model [54] synthesizes sequences of hard impact events of short duration in order to simulate those sound sources whose emission is interpreted by the hearing system as a superposition of microscopic crumpling events. Such a model is based on the impact model on top of which a statistics of temporal impact events is superimposed. In addition to the control parameters inherited from the impact model, which define the atomic events, the model is controlled by the force of the crumpling process and the crumpling resistance.

3.2.5. Solid–liquid interaction model

This model was developed adapting previous works on fluid simulations [55–57] to the case of footstep sounds. As opposed to those studies, the proposed model relied on a physically inspired approach which did not involve a fluid simulator based on graphic solutions (such as the shallow water model or the smoothed particle hydrodynamics). To emulate the sound of a footstep on a liquid volume (e.g., a puddle of water), the physical processes involved in the sound generation caused by the impact of a rigid body with a fluid were simulated. Specifically, the designed model was inspired by research on the sound generation process of a projectile impacting and entering a liquid volume. According to Richardson [58], such a process consists of three phases: (i) the initial impact; (ii); the generation of a set of small bubbles; (iii) the generation of a large bubble simulating the oscillation of the air cavity produced by the complete penetration of the rigid object into the fluid volume. Nevertheless, in the case of the foot entering in the liquid volume, the large bubble is not generated since the water does not seal any air cavity forming at the surface. Therefore, the phase (iii) was not simulated.

In previous research [56,57] the sound produced by the initial impact of the solid object with the liquid was simulated by means of a resonant filter [42] producing a brief impulsive noise. However, in the case of the footstep sounds, such an impact is a continuous process rather than a discrete event. Therefore, the impact of the foot with the liquid volume was rendered by multiplicating the exciter signal defining the foot–floor interaction with a low-pass filtered white noise. This solution was in agreement with Richardson's observations concerning the proportionality of the intensity of the impact sound between a rigid body and a fluid with the speed of the body at the moment of impact.

To generate the set of small bubbles, a single resonating spherical bubble model was utilized to serve as building block for a fluid simulator [59,60]. The equation for the pressure wave created by an oscillating spherical bubble is given by:

$$p(t) = A sin(2\pi f(t)t)e^{-dt}$$

where the amplitude A is determined by the initial excitation of the bubble (according to Longuet-Higgins [61] $A = \epsilon r$, with $\epsilon \in [0.01, 0.1]$ and r being the bubble radius); the time-varying resonance frequency is given by $f(t) = f_0(1 + \xi dt)$, with $f_0 = 3/r$ (Minnaert's formula [62]), $\xi = 0.1$ for water according to Van Den Doel's observations [60], and d a damping factor (for water $d = 0.13/r + 0.0072r^{-3/2}$, denser liquids are characterized by greater damping factors [61]).

The simulation of the small bubbles population was controlled by a fluid simulator based on a bubble generation criteria determining how many bubbles were present at a given instant of the foot-fluid interaction, and a stochastic model defining the sound of each bubble. Specifically, the amount of bubbles was determined by the combination of two factors used as tuneable input parameters for the model: the liquid velocity and the depth of the liquid volume. To emulate the liquid velocity the exciter signal was utilized as input for a delay line (20 ms) with feedback. The amplitude of the delayed signal was scaled in order to regulate its temporal duration. The resulting signal was then multiplied by a factor expressing the depth of the liquid volume. The idea underlying such a design choice was that deeper liquid volumes excited by stronger foot interactions generate greater amount of bubbles, and as a consequence the produced sound lasts longer.

The sound of each bubble was determined by several parameters. Following the approach adopted by Moss et al. [55], the radius of each bubble r and the initial excitation ϵ were selected at random according to a physically inspired approach based on power laws: $r^{-\alpha}$ and $\epsilon^{-\beta}$, having α as tuneable parameter ranging between 1.5 and 3.3 and $\beta = 2$. In addition, the proposed model accounted also for the tuning of the minimum and maximum generable radii (typically in the range [0.15,10] mm). In particular, the maximum value was related to the parameter controlling the depth of the liquid, following the design choice of producing bubbles with bigger radii (and therefore with lower resonance frequencies [62]) for deeper liquid volumes. Following the same approach, the depth of the liquid was also related to the value of the cut off frequency of the low pass filter used to model the initial impact. Finally, the model accounted for the control of the liquid density, expressed as a tuneable parameter $\rho \geqslant 1$ multiplying the damping factor described above for water.

3.2.6. Excitation and combination the sound models

The exciter signals were used to control some of the parameters of the models described above along with the range of variation of the amplitudes of both the independently simulated subcomponents and the global sound. Such a control was either "direct", i.e., the whole signal was used as input for a parameter of the model, or "indirect", i.e., some features of the exciter were extracted (e.g., its maximum value, beginning, end) and coupled with random calculations or various types of functions (e.g., linear ramps).

Specifically, the exciter directly controlled the system energy parameter in the PhISM, the impact velocity in the impact model controlled "in velocity", the force in the impact model controlled "in force", and the profile envelope of the fractal noise model; it indirectly controlled the external rubbing force and the exerted pressure on rubber in the friction model as well as the force of the crumpling process in the crumpling model, and it was used as basis of the fluid simulator for the liquid model (see Section 3.2.5).

As a consequence of the design paradigm based on the independent rendering and subsequent merging of the sound subcomponents, the models were combined together most of the times in order to reproduce the wanted global sound. Firstly, some ground materials were synthesized by using multiple instances of the same model. An example is the sound of different types of gravel where two or three instances of the PhISM were utilized to sonically reproduce two or three kinds of distinct collisions between stones of the same dimension. Secondly, some ground materials were simulated by means of different models. For instance, to simulate footsteps on a creaking parquet, the impact model was used to render the wood-shoe contact, while the friction model was utilized for the creaking component; another example is the sound of snow, where one PhISM and one crumpling model were coupled in order to simulate the two subcomponents of the sound produced when the foot drops into the snow and the snow breaks under the foot respectively. Also, to model hybrid surfaces such as asphalt with some small pebbles, the impact model was utilized in combination with the PhISM. In addition, to simulate wet surfaces, such as wet gravel, the sound produced to render the corresponding dry surface was summed to its multiplication with the liquid model. In this case, however, the noisy sound emulating the initial impact in the liquid model was not produced, since a liquid volume is not present and the resulting physical process is different. Table 1 illustrates a schematic summary of the techniques utilized for the synthesis of the various typologies of ground materials.

Regarding the modeling of the different types of foot–floor interaction illustrated in Section 3.1, the exciter was utilized as described above to drive the physical models simulating the various surface materials. However, for the case of the solid surfaces, the sliding and brushing interactions were rendered differently from the aggregate and liquid surface typologies. Specifically, in presence of a sliding interaction on a solid surface, the impact model was utilized in combination with the fractal noise model, while the brushing interaction was simulated using only the fractal noise model. Table 2 shows a schematic summary of the techniques utilized for the synthesis of the various types of foot–floor interaction according to the different surface typologies.

3.3. Modeling the shoe contribution

The most important structural invariant related to the shoe type is the hardness of its sole [22]. Stereotypically gendered shoes, such as high heels for females and dress shoes for males are characterized by hard soles compared to genderless shoes, such as sneakers with soft rubber soles. However, the everyday experience suggests that perceptual relevance of the sole hardness is greater for the solid surfaces compared to the aggregate ones, and scarcely for the liquid ones. Therefore, the sole hardness was modeled

Table 1
Techniques utilized for the ground materials synthesis.

Ground typology	Technique
Solid (dry)	I
Solid (wet)	I + I*L
Solid plus aggregate	I + P
Solid plus friction	I + F
Aggregate (dry)	P[3], P[3] + C
Aggregate (wet)	P[3] + P[3]*L, P[3] + C + (P[3] + C)*L
Liquid	L[3]

I = impact model, P = PhISM, F = friction model, C = crumpling model, L = solid–liquid interaction model, [3] = up to 3 instances of the model, + = additive synthesis, * = signals multiplication.

Table 2
Techniques utilized for the foot–floor interaction synthesis.

Interaction type	Technique
Normal step	S(E), A(E), L(E)
Running step	S(E), A(E), L(E)
Sliding	FN(E) + S(E), A(E), L(E)
Brushing	FN(E), A(E), L(E)

S, A, and L = techniques shown in Table 1 for solids, aggregates and liquids respectively, FN = fractal noise model, (E) = exciter signal in input for the model, + = additive synthesis.

thinking mainly to its effect on the sound resulting from the interaction with a solid surface, although it was conceived to be still used while exciting aggregate and liquid materials.

The distinctive clicking sounds produced by stereotypically gendered shoes on a solid surface were simulated by creating a sharp exciter signal with short attack and decay for both heel and toe (see Fig. 2(a) and (b)). Conversely, the exciter simulating genderless shoes with soft soles was characterized by the smoothness of the attack and peak values generally lower (see Fig. 2(d)). On the one hand, this choice was inspired by the findings reported by Freed [63], which showed that the hardness of a mallet impacting metal objects can be predicted only by the acoustical information contained in the type of attack. On the other hand, it was inspired by the common everyday experience: softer soles produce sounds having amplitude lower than harder ones.

Moreover, in the impact model controlled “in velocity”, the sole hardness was also defined by the k, λ, α and the mass of the hammer parameters. In the proposed simulations, lower values of force stiffness, energy dissipation and hammer mass, and higher values of the contact surface shape were associated with softer soles, while the opposite for the hard ones. This choice partially paralleled the findings reported by Giordano and Petrini [64] for the perceived hardness of a mallet, which was modeled by the force stiffness coefficient k. However, better simulations of boots and sneakers were found using the impact model controlled was “in force”, while dress shoes and high heels were always simulated with the impact model controlled “in velocity”. In addition, the sound of sneakers on a solid surface is typically characterized by a noisy component. To achieve such a distinctive feature, the signal of the impact model was added to its version multiplied by a white noise signal.

These manipulations of the temporal evolution of the exciter and of impact model’s parameters in order to simulate different levels of sole hardness, resulted, for the same simulated surface material, in sounds having different properties in terms of attack, peak, and duration, as shown in Fig. 3.

Other structural invariants of the shoe material eventually contributing to the footstep sound can be simulated by adding a further sound-layer. This is for instance the case of the squeaking sometimes produced by certain types of shoes (e.g., training shoes or new dress shoes). Such a squeaking component was simulated by using the friction model. Another example is the typical sound of the metallic buckle or spurs present in some types of boots. Such a sound was rendered by means of a bouncing model which simulates the falling of an object against another (details can be found in [65]). Also, some types of shoes (e.g., some slippers) are characterized by “microcliks”. Such low amplitude sounds were simulated by means of the PhISM. Table 3 shows a summary of the techniques utilized for the synthesis of the various shoes sound components.

3.4. Modeling the walker

Two anthropomorphic features of the walker were simulated, namely the body size (both height and weight) and foot-length. According to Li et al. [19] such features are related to the auditory perception of the gender. Specifically, they found that the anthropomorphic feature most correlated with maleness judgments was the height. Nevertheless, they suggested that the gender identification is determined by the center of mass (which accounts for the combination of height and weight). The findings reported by Li et al. [19], which were confirmed by those reported by Giordano and Bresin [22] and by Visell et al. [21], revealed that when significant body size differences between males and females are present, the gender perception is related to spectral properties of the footstep sounds: sounds having spectra with a predominant high frequency component were associated with females (i.e., small body sizes), while maleness (i.e., big body sizes) was related to spectral dominancy of the low frequencies. In addition to this, bigger body sizes are expected to produce louder sounds since the heavier and taller the person, the greater the GRF.

Therefore, the synthesis of the person’s body size, which in turn is related to the gender, was achieved by manipulating both sound’s amplitude and spectrum (i.e., varying both the position of the spectral mode and the amount of low or high frequencies). Specifically, five body sizes were simulated: big, medium-big, medium, medium-small, and small. Therefore, these can also be interpreted in descending order of maleness, being the medium body size considered as genderless. The synthesis process consisted of the alteration of the spectrum and the amplitude of the sound corresponding to the medium body size: bigger/smaller body size were rendered by adding to the original sound a low-pass/high pass filtered version (appropriately scaled in volume) and by using a high-shelf/low-shelf filter to attenuate the high/low frequencies, as well as by involving an exciter with greater/smaller peak. In presence of solid surfaces rendered with the impact model controlled “in velocity”, greater/smaller values of the mass parameter were utilized. Importantly, such alterations were appropriately made in order to keep unaltered the perception of the type of material and shoe involved.

In a different vein, the foot-length was rendered, in non-interactive contexts, by the step’s temporal duration, such that, having equal walking speed (and type of shoe), the longer the foot the greater the duration. To achieve this, the parameters governing the temporal evolution of the exciter were appropriately tuned (in first place, the temporal distance between heel and toe strikes). This design choice was motivated by findings reported by Li et al. [19] about the differences between the step durations of males and females. Five foot-lengths were rendered, one for each of the five body sizes simulated. Furthermore, in the proposed simulations, the duration of each step was also dependent on the speed of walking. In general, it is worthwhile to notice that the everyday common experience suggests that the perception of a walker’s gender can also be related to the type of shoes worn (e.g., the sound of high heels is more likely to be associated with a woman rather than a man).

4. Sound models control

4.1. Modeling the locomotion

A sequencer implementing a simple biomechanically inspired model was developed to control the production of footstep sounds for the non-interactive simulation of different types of locomotion, such as walking, running, sprinting, running decelerations, and jumping in place. To this end, the different exciter types for walking, running and jumping described in Section 3.1 were involved.

The walking and running cycle is globally represented by three parameters: the step length (l), the step frequency (f) and the walking/running speed (v), which are linked through the relation: $v = l * f$, where $f = 1/T, T$ being the temporal distance between

Table 3
Techniques utilized for the shoes synthesis.

Component	Technique
Hard sole type (in solids)	$I(k, \alpha, \lambda, m)$
Medium-hard sole type (in solids)	E or $I(k, \alpha, \lambda, m)$
Soft sole type (in solids)	E*N
All sole types (in aggregates)	E
All sole types (in liquids)	E
Frictions	F
Microclicks	P
Buckles	B

E = exciter signal, $I(k, \alpha, \lambda, m)$ = parameters controlling the impact model, F = friction model, P = PhISM, B = bouncing model, N = white noise, * = signals multiplication.

adjacent steps [66]. At a given speed, one can walk or run with infinite combinations of l and f. In general, however, in order to increase walking or running speed, humans jointly increase l and f [67]. At auditory level, however, the fundamental parameter to recognize v is only T, while l plays a minor role [68,31]. In addition, according to the results reported in [43,67], v is related to the GRF such that the higher the v the greater the GRF's amplitude and the shorter the GRF's duration. Therefore, T was used for rendering v, and its values were used to control both amplitude and the duration (i.e., the temporal distance between heel and toe strikes) of the exciter corresponding to each step. In addition, the simulations were based on the works of Nilsson and Thorstensson [43,67], who calculated values for both the minimum T (about 210 ms for walkers and 140 ms for runners), and for the maximum temporal distance between heel and toe strikes (150 ms).

As far as the jumping action is concerned, only one control parameter was used, i.e., the height of each jump. It directly controlled both T (the temporal distance between consecutive jumps) and the GRF's amplitude, such that the greater the jumps' height the greater T and the exciter's amplitude. In general, the amplitude of the exciter was different for the three types of locomotion associated with a same person (see Fig. 2): indeed the GRF's amplitude is greater for jumping than for running, and smallest for walking [43–45]. In addition, since the GRF is related to the body weight, the exciter amplitude was also adjusted taking into account the simulated body size (see Section 3.4).

On a separate note, the proposed sequencer took into account the fact that in real life, the sound of each step is different from the previous one. The results presented in [29] demonstrated that the concatenation of the same footstep sound in sequences of walking and running sounds is perceived as mechanical. Therefore, to increase the perceived realism of the different types of locomotions, each step in a sequence was rendered in a different way. On the one hand, this was achieved by generating a different type of exciter for each step (e.g., using different times of attack and decay or different peak values for heel and toe, choosing the amount of steps in a sequence having the toe component, controlling the density of the brushing events between two subsequent steps); on the other hand, by calculating for each step a different set of appropriate values for the parameters controlling the models.

4.2. Tuning the sound models

One of the challenges in the synthesis was to find the suitable combinations of parameters and their range of variations that could provide a perceptually convincing simulation. Following the tenets of the analysis-by-synthesis technique, on the one hand the values of all the control parameters were tuned empirically until the sonic result was perceptually in agreement with the average everyday experience of the wanted simulation; on the other hand, such a tuning, was heavily based on the knowledge produced by sound perception research, as discussed below. In particular, the findings of perceptual experiments validating the previous version of the synthesizer [28] were taken into consideration. Furthermore, during the whole design process the tuning of the synthesis algorithms was subjected to extensive preliminary listening tests to assess the perceptual validity of each assumption.

Each ground material was designed by impacting the resonator with an exciter corresponding to the heel of the hard sole of a high heel shoe (see Section 3.3). This approach was chosen in order to achieve the punctual excitation of a hard hammer which can better reveal the intrinsic properties of a material (especially for solid surfaces). Also, sounds were designed in absence of any reverberation, like if they were produced in an anechoic setting, considering the fact that the addition of reverberation would affect their spectral content. Moreover, the ground materials were designed by considering a genderless walker (i.e., with a medium body size), so that the spectral content of the sounds could be changed according to the person's anthropomorphic features (see Section 3.4). In addition, since physical models were involved, widely available tables of material parameters specifying some of the mechanical properties of the sound sources (e.g., stiffness for solids, grain size and compliance for aggregates) were taken into account to adjust the synthesis parameters. These tables, however, do not provide for each material a unique value of the considered property, but a range (for instance, different types of wood exist) and sometimes such a range overlaps between materials.

As far as the solid surfaces are concerned, the tuning of the impact model leveraged the results of investigations on sonic interactions between objects in contact, which describe the relationship between physical and perceptual parameters (for a recent review see [40]). According to several studies, the material perception of a real or virtual object impacted with a real or virtual hammer is mainly correlated with the damping of spectral components and that such a decay-time plays a much larger role than frequency [69–72,64,73]. Overall, studies on the identification of impacted solid materials revealed a nearly perfect ability of listeners to distinguish between gross material categories (e.g., wood or plastic vs. metals or glass), while the material identification within these categories depends on the sound's spectral color [40]. As far as footstep sounds are concerned, Fontana et al. showed that spectral color is an important parameter in the recognition of walked-upon solid surfaces [11]. Research has shown that the sound's duration and spectral color resulting from solid object collisions depend on the hardness properties of both the hammer and the sounding object, and that these are confused at perceptual level in scaling and identification tasks [64,39]. Overall, research has shown that the perception of the material properties of the sounding object is influenced by both the properties of the sound decay and sound frequency, while the perception of the material and mass of the hammer is influenced by both loudness and spectral centroid [40]. Stiffer sounding object's materials produce higher frequency spectral components and spectral components characterized by a slower decay, while stiffer hammer materials produce an increase in the high-frequency energy of the radiated sound.

In the proposed simulations, the parameters of the sounding object in the impact model were tuned according to available physical measures of the level of hardness of the solid material in hand (precisely, tables with values of stiffness of various materials). Such a mechanical material property is linked to both sound decay and frequency [40]. Mainly leveraging the results reported by Giordano et al. [64,39], firstly the decay parameter for each mode utilized in the impact model was tuned according to the wanted material hardness: high decay values were set for hard or stiff materials (e.g., steel), while low decay values were set for soft or elastic materials (e.g., rubber). Secondly, harder materials were rendered by tuning the modes' frequencies in order to have a spectral centroid placed at higher frequencies than softer materials. Such a tuning was furthermore inspired by the spectral analysis of different recordings of footstep sounds on solid materials aiming at finding for each solid material a range of frequencies that could characterize it independently from the shoe type and the room reverberation effect. In tuning the decay parameters of each mode, it was taken into account the fact that the damping is frequency dependent (high frequency components are more rapidly damped than low-frequency components) [40].

Considerations about the tuning of the sound models for simulating the shoe types are presented in Section 3.3.

Regarding the simulation of aggregates and liquids, a followed guideline was that greater objects generally have a lower fundamental frequency than smaller objects [20]. Therefore, for instance, gravels with big grain sizes were modeled by tuning the PhISM's pole radius and frequency parameters (see Section 3.2.1) in such a way that the resulting particles frequency was lower than that of the particles simulating gravels with small grain sizes. Similarly, the tuning of the depth of a liquid material was based on the guideline of having not only a greater number of bubbles but also an amount of bubbles with bigger radii for greater depths (see Section 3.2.5). Another guideline was that the compliance of an aggregate material, which is related to the density of the grains, was controlled by tuning the PhISM in order to achieve an appropriate amount of colliding particles and sound duration. Analogous guidelines were utilized for the tuning of the impact events of the crumpling model.

In regards to the additional sound layers to model some structural invariants of the shoe material (see Section 3.3), the tuning of the bouncing model simulating the buckles/spurs followed again the guideline inspired by results of Coward and Stevens [20] (bigger buckles/spurs having lower fundamental frequency); as far as the microclicks are concerned, the amount of colliding particles parameter in the PhISM was tuned in order to render the wanted amount of microclicks.

It is important to notice that when modeling the person's body size, by adding to the simulation of the medium body size (i.e., genderless person) a filtered version to enhance low or high frequency content (see Section 3.4), the tuning of the spectral filters was done in order to keep unaltered the perception of the type of material and shoe involved. The amount of added high or low frequencies as well as their amplitude, varied for each material and type of shoe.

In general, all the tunings described above on the one hand followed the results of sound perception research, on the other hand they were contingent to the author's design choices, which were anyhow based on common everyday experience. The latter were carefully tested during the whole process, in agreement with the tenets of the analysis-by-synthesis method. The full evaluation of all the proposed techniques and design choices is reported in Section 5.

4.3. Intuitive control

In order to allow sound designers to create the footstep sounds that they have in their mind, a control strategy of the large number of the synthesis parameters is necessary. Here it is described one of the possible strategies that provides an intuitive control of the synthesis parameters based on the evocations of sound sources for the class of footstep sounds. It has been inspired by the work presented by Aramaki et al. [74].

The proposed control strategy is based on three hierarchical layers (see Fig. 4). The first layer offers the most intuitive way for a non-expert user to create footstep sounds: the control is based on verbal descriptions of the mental representation of the sound source, like the type of foot–floor interaction (e.g., walking step, sliding), ground material (e.g., wood, snow), shoe (e.g., boots, sneakers), locomotion (e.g., fast running, high jumping), and person performing it (e.g., male, big body size). The second layer represents the controls that an expert user will have to manipulate in order to define the subcomponents of the wanted footstep sounds relevant from a perceptual point of view, i.e., which invariants of the sound source need to be simulated. At this stage the user decides which and how many subcomponents to enable, how they are combined (i.e., by additive synthesis or signals multiplication) and the volume with which each of them contribute to the global sound. For each subcomponent a set of presets of the synthesis parameters is provided. The third layer is composed of the sound models described in Section 3.3 that define each sound subcomponent.

The proposed high-level control strategy (first layer) offers various possibilities of sound creation and of sound effects based on few intuitive control parameters. For example, a user can easily generate sounds of walking on the same ground produced by various walkers wearing different shoes. Table 4 summarizes the currently implemented verbal descriptors and their values.

To achieve coherence and consistency, not all the descriptors' values are available at any moment. For instance, when a user select the genderless value for the gender descriptor, the available body-size value is only medium (see Section 3.4); or, when the jumping locomotion type is selected then only the jumping step foot–floor interaction is available among the values of the steps descriptor. It is worthwhile to notice that by default, the user only has access to the top layer. Nevertheless, an expert user is given the possibility to directly access the other two layers.

Two mappings were implemented between these three layers (represented as black arrows in Fig. 4). As the parameters that allow intuitive controls are not independent and might be linked to several signal characteristics at a time, the mappings are far from being straight-forward. The mapping between the first and second layers (i.e., from sound source to sound source subcomponents) was based on the results of the phase of analysis (see Section 2.3) in which the invariants of the footstep sound sources were identified and considered as subcomponents of the sound. As far as the foot–floor interactions are concerned, the types of steps (e.g., walking, running, or jumping steps) were mapped to the impact transformational invariant, while the other types of foot–floor interactions (e.g., sliding, brushing) were mapped to the friction transformational invariant. The types of shoes were mapped into the structural invariants sole hardness (e.g., dress shoes), squeakiness (e.g., training shoes), clickiness (e.g., slippers), and metallic (e.g., boots with buckles/spurs) according to their properties. As far as the surface materials are concerned, solids were mapped to the structural invariants stiffness and creakiness; aggregates to homogeneity, compliance, granularity, and fracture; liquids to viscosity, density, and depth; hybrid materials were mapped accordingly. Regarding the modeling of the person, gender and anthropomorphic features were mapped to the structural invariants foot-length and body size. The different types of locomotion were mapped to the transformational invariants impact, friction, time between steps, and step amplitude.

The mapping between the second and third layers (i.e., from sound source subcomponents to sound models) was defined by assigning to each subcomponent a sound model capable of simulating it. The subcomponents impact and friction were associated with the exciter model; sole hardness with both the exciter model and the impact model; squeakiness, clickiness, and metallic with friction model, PhISM, and bouncing model respectively; stiffness was rendered with the impact model, creakiness with the friction model; granularity with the PhISM; homogeneity, compliance, and fracture with both the crumpling model and the PhISM;

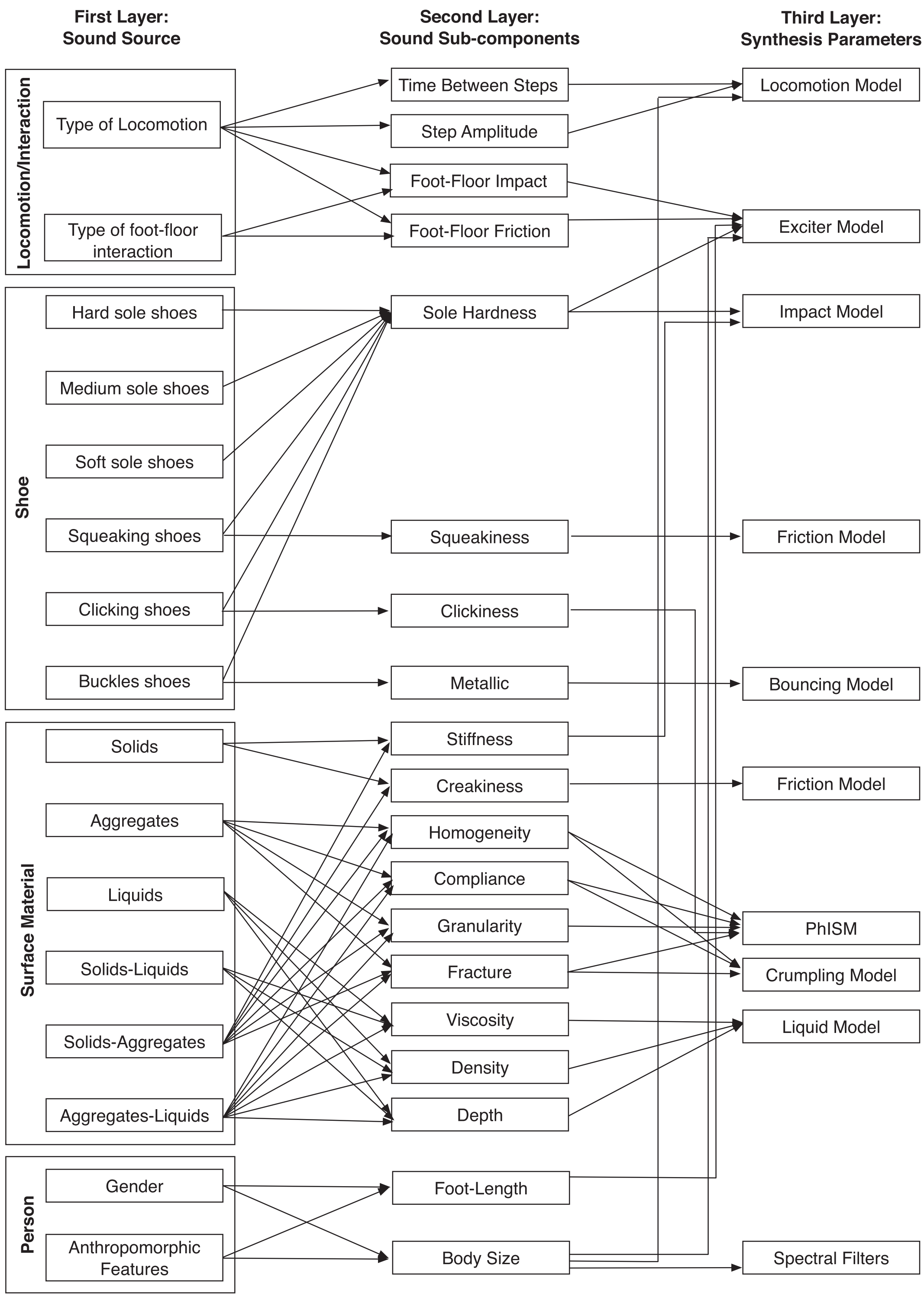

Fig. 4. The three-layers control strategy.

viscosity with the solid–liquid interaction model; foot-length was mapped to the exciter model, while body size to both the walker model and the exciter model. Heel-to-heel and step amplitude were mapped to the locomotion model presented in Section 4.1.

The analysis-by-synthesis technique was then adopted to validate the proposed mappings, as well as to determine the synthesis parameters of each sound model. Such a validation is reported in Section 5.

4.4. Implementation

Using the algorithms and the sound design paradigms described in previous sections, a comprehensive collection of footstep sounds were implemented.[2] Specifically, 8 types of shoes, 40 types of surface materials, 5 types of person were simulated, which combined

[2] Audio excerpts of the described simulations can be downloaded at www.ahws-project.net.

Table 4
List of intuitive controls based on verbal descriptions.

Descriptors	Available values
Locomotion	
Type of locomotion	Walking, walking with scuffs, running, jumping, sprinting, decelerating, limping, sliding, scuffing
Shoe	
Hard sole shoes	Dress shoes, high heels
Medium sole shoes	Boots
Soft sole shoes	Sneakers
Squeaking shoes	Squeaking dess shoes, squeaking sneakers
Clicking shoes	Clicking dress shoes
Buckles shoes	boots with buckles/spurs
Surface material	
Solids	Wood (3 types), creaking wood (6 types), concrete (2 types), metal (2 types), marble (2 types)
Aggregates	Leaves, dry leaves, sand, soft deep snow, crunchy snow, dirt, grass, forest underbrush, gravel, coarse gravel, fine gravel
Liquids	Water puddle, oil puddle (both with low, medium, and high depth)
Hybrids: solids–liquids	Wet concrete
Hybrids: solids–aggregates	Concrete plus small pebbles
Hybrids: aggregates–liquids	Mud, wet gravel, wet coarse gravel, wet fine gravel, wet sand, wet forest underbrush
Person	
Gender	Male, female, genderless
Anthropomorphic features	Foot-length, body size (both with big, medium-big, medium, medium-small, and small values)

between each other give rise to a total of 1600 possible footstep sounds. These in turn can be organized in 9 types of locomotion. Table 4 shows all the implemented simulations. Furthermore, designers, by acting on the second and third layers of the adopted control strategy (see Section 4.5), can create their own footstep sounds, giving rise to limitless possibilities.

The footstep sounds synthesizer was developed under Max/MSP[3] and Pure Data[4] sound synthesis and multimedia real-time platforms. It runs over mac, windows and linux operative systems. The computational cost of the algorithms is rather low: they run easily on ordinary computers and even mobile phones or other portable systems (e.g., [75]). In more detail, the implementations of the models for impact, friction, crumpling, and bouncing present in the Sound Design Toolkit [49] were utilized. The PhISM, the fractal noise model, and the solid–liquid interaction model were implemented in C++ as external libraries.

On the other hand, the signals defining the different types of exciter were implemented using either the facilities of Max/MSP or MATLAB.[5] In the latter case, each signal was defined by a function varying between 0 and 1 on a grid of 44,100 points per second, created by the spline interpolation of a certain number of points set to achieve the wanted foot–floor interaction. For instance, the part of the exciter corresponding to the heel in a normal step was defined by two splines of 11 points each, to define the attack and the decay respectively (see Fig. 2(a)). As far as the sliding and brushing are concerned, the corresponding exciters were created by building a signal by means of spline functions and altering each sample at random within an appropriate range (see Fig. 2(i) and (j)). Except when using MATLAB to create the exciter signal, all the manipulations of the synthesis parameters can be done in real-time. In this way the developed engine serves as a tool to explore the sound space of the foot–floor interactions in an interactive manner.

The synthesized sounds were also enhanced with reverberation algorithms based on the technique of convolving a signal with an impulse response corresponding to a room with specific dimensions. Specifically, such an approach was possible in real time, thanks to the tools for Max/MSP allowing convolution with zero latency developed by Harker and Tremblay [76].

[3] www.cycling74.com.
[4] www.puredata.info.
[5] www.mathworks.com.

4.5. Control in interactive and non-interactive scenarios

The developed synthesizer can be utilized in both in interactive contexts (where users perform physically the locomotion) and non-interactive scenarios (where the locomotion is simulated while users are sitting on a chair). The control of the synthesizer in the interactive case is achieved by means of locomotion interfaces, such as instrumented shoes [75,18] or augmented floors [77,18]. While the non interactive case has no limitations on the use of all the developed simulation features, when passing to technologically-mediated foot–floor interactions, the control of the synthesis algorithms needs to be adapted to the affordances of the locomotion interface at hand [78].

The current locomotion interfaces usable to control interactively the synthesizer can be divided in two categories: those based on a system of microphones for the extraction of the exciter signal from footstep sounds and those based on the triggering of ad-hoc built exciter signals [18]. In the first case the foot–floor interaction and the type of shoes are already fully defined by the extracted exciter, which, therefore, is used only to drive the models for the simulation of the ground. Conversely, in the second case, it is possible to tune the synthesizer to simulate a type of shoes different from those worn by the walker, as well as a type of interaction different from that actually produced by the foot. In both the cases, however, it is possible to manipulate the spectral content of the sounds in order to render the walker's body size properties (and as a consequence the gender, see Section 3.4). Figs. 5 and 6 illustrate a schematic representation of the developed architecture for interactive and non interactive scenarios respectively.

5. Evaluation

The evaluation consisted of four listening tests intended to assess the perceived plausibility of the synthesized sounds: (i) locomotions identification, (ii) surface materials identification, (iii) shoes identification, (iv) anthropomorphic features identification.

5.1. Participants

Fourty-eight participants (24 M, 24 F), aged between 21 and 34 (mean = 25.7, SD = 3.23), were divided into four groups ($n = 12$) to perform the four experiments. All participants reported normal

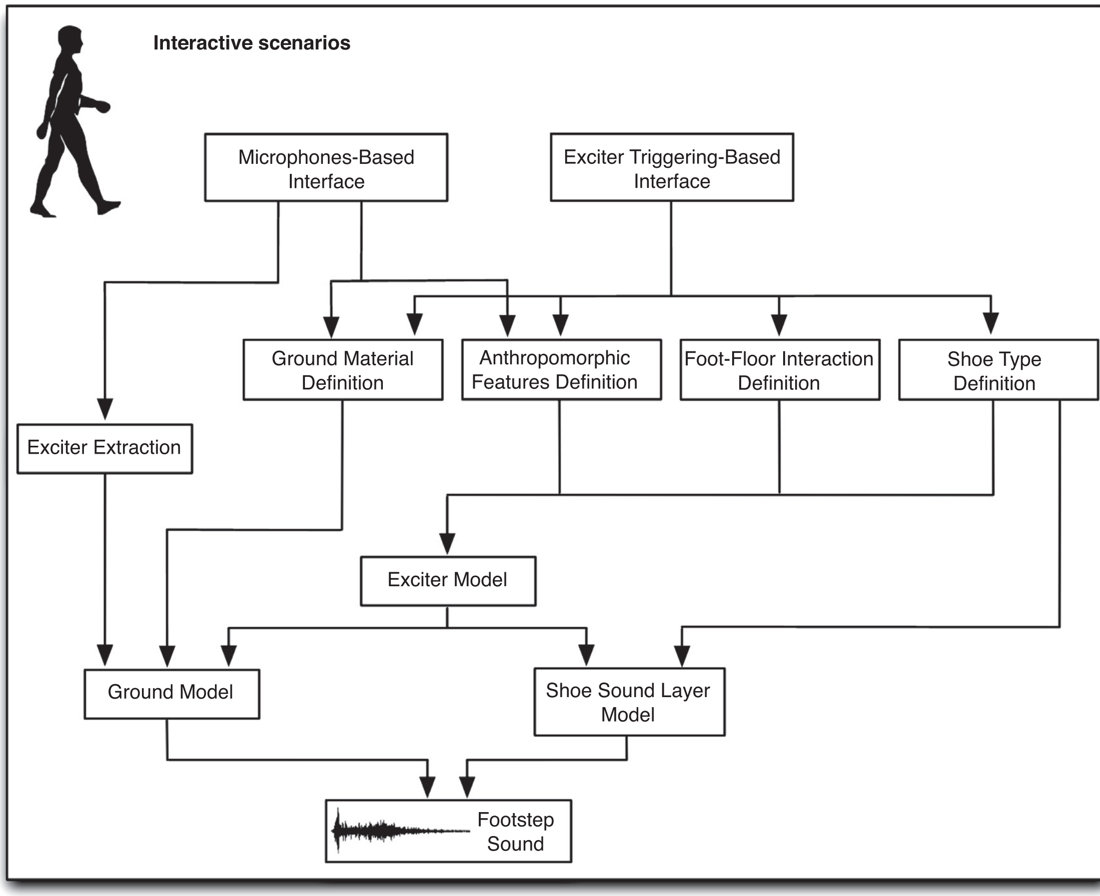

Fig. 5. Schematic representation of the developed architecture for interactive scenarios.

hearing conditions. They took 18, 57, 35, and 20 min to complete experiment 1, 2, 3, and 4 respectively.

5.2. Apparatus

The setup of the experiment was installed in a silent room and consisted of a laptop (Macbook Pro), a soundcard (Fireface UFX), and a closed headphone set (Sennheiser PXC 450). The laptop run the footstep sound synthesizer described in previous sections as well as a graphical user interface, both realized in Max/MSP. The latter consisted of buttons to start and stop the trials and a list of labels displaying the possible choices for the identification task, which were coupled with as many visual analog scales (VAS) in the range [not plausible at all, very plausible]. Each of the labels represented one of the possible occurring stimuli. However, participants were not informed that each of these choices would have appeared as stimulus.

5.3. Procedure

Participants were presented with written instructions. They were asked to wear the headphones set and were instructed to start each trial by pressing the corresponding buttons of the graphical user interface. After the presentation of the stimulus, participants were asked to rate its plausibility on each of the VAS coupled with the possible choices in the list. In more detail, in experiment 1, where the plausibility of stimuli representing different types of locomotion was assessed, the question was: *"To what extent it is plausible that the sounds you listened to represent the following types of locomotions?"*. In experiment 2, where the plausibility of stimuli representing different types of surface materials was assessed, the question was: *"To what extent it is plausible that the sounds you listened to represent the following types of surface materials?"*. In experiment 3, where the plausibility of stimuli representing different types of shoes was assessed, the question was: *"To what extent it is plausible that the sounds you listened to represent the following types of shoes?"*. In experiment 4, where the plausibility of stimuli representing different types of walker's anthropomorphic features was assessed, the questions were: *"To what extent it is plausible that the sounds you listened to represent the following types of body size?"*; *"To what extent it is plausible that the sounds you listened to represent a female walker?"*.

By asking to rate the plausibility of the stimuli on a list of continuous scales coupled with a label indicating a possible choice, not only allowed to assess the validity of the design choices made for each stimulus, but also to identify which of the labels on the list was rated as the most plausible as well as which labels were evaluated among the less plausible.

In each experiment, stimuli were repeated twice and presented in randomized order. When activated, each stimulus was looped with an interval of 3 s between the repetitions, so participants could listen to it as much as they needed before giving the response. In this way participants were provided with the best possible conditions for the use of the auditory information carried by the stimuli: confidence with the response was preferred both to the control of participants' response time and exposure to stimuli. When the answer was chosen, the sound stopped and all the sliders of the interface were automatically set to the minimum value. When passed to the next stimulus participants could not change their response to the previous stimuli. Before the beginning of the experiment, they familiarized with the setup and with the identification procedure practicing with excerpts not included in the experiment.

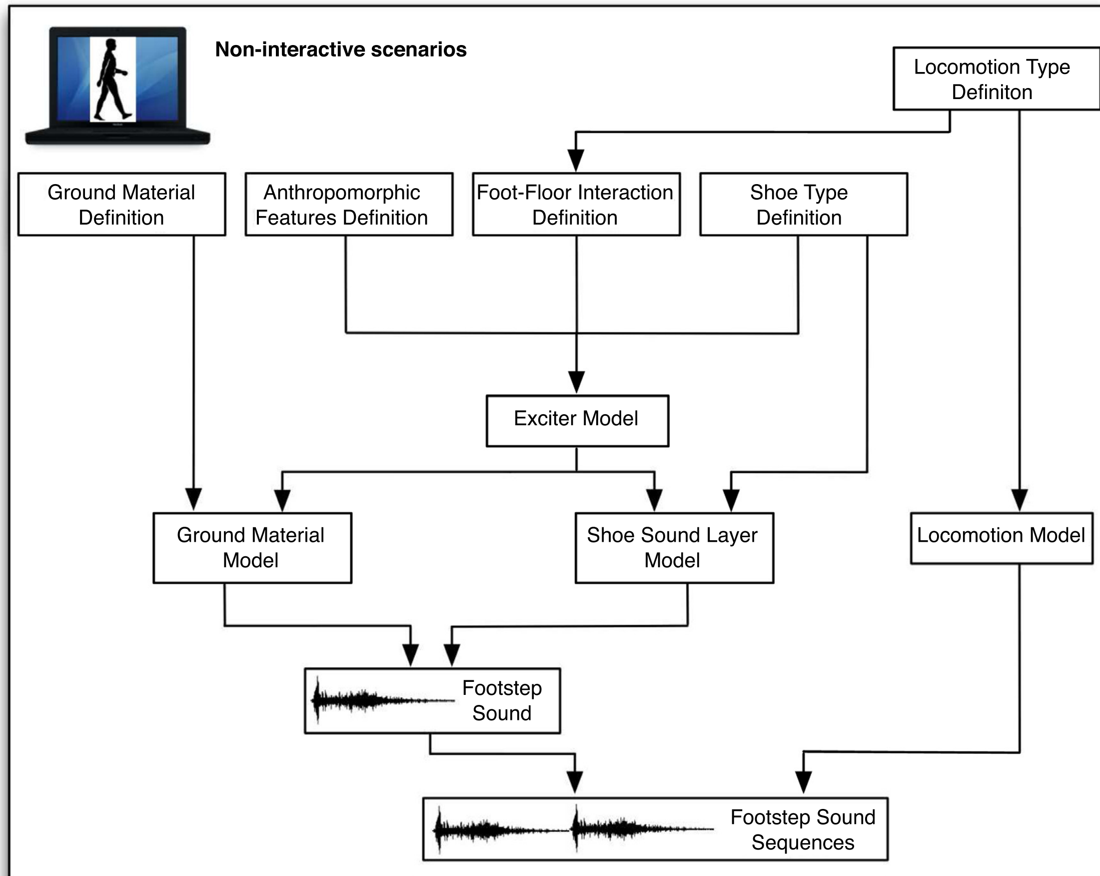

Fig. 6. Schematic representation of the developed architecture for non interactive scenarios.

5.4. Stimuli

The stimuli of Experiment 1 consisted of the simulation of 5 locomotion types (walking, walking with scuffs, running, jumping in place, and sliding), performed by a genderless walker (i.e., with medium body size), on 2 surface materials (wood and gravel) wearing two types of shoes (genderless dress shoes and sneakers). The resulting 20 stimuli were repeated twice for a total 40 trials. Specifically, the locomotion model described in Section 4.1 was set to have a T equal to 800 ms, 1000 ms, 300 ms, 650 ms, and 2000 ms for walking, walking with scuffs, running, jumping in place, and sliding respectively. Each trial was composed by a sequence of 6 steps. The exciters corresponding to the steps in each locomotion type were selected according to the techniques described in Section 3.1. The familiarization phase involved five stimuli simulating the five locomotion types performed on mud by a genderless walker wearing boots.

The stimuli of Experiment 2 consisted of the simulation of walking (T = 800 ms), performed by a genderless walker (i.e., with medium body size) wearing two types of shoes (genderless dress shoes and sneakers), on the 21 surface materials listed in Tables 6 and 7. The resulting 42 stimuli were repeated twice for a total 84 trials. Each trial was composed by a sequence of 6 steps. The familiarization phase involved five stimuli simulating walking performed by a genderless walker wearing boots, on hardwood, wet sand, and high grass.

The stimuli of Experiment 3 consisted of the simulation of walking (T = 800 ms), performed by a male and a female walker (i.e., with big and small body size respectively) wearing the seven types of shoes listed in Table 8, on three surface materials (Wood, Gravel, and Water).The resulting 33 stimuli were repeated twice for a total 66 trials. Each trial was composed by a sequence of 6 steps. The familiarization phase involved seven stimuli simulating the seven types of shoes performed on mud by a genderless walker.

The stimuli of Experiment 4 consisted of the simulation of walking (T = 800 ms), performed by walkers having three body sizes (big, medium, and small) wearing two types of shoes (genderless dress shoes and sneakers), on three surface materials (Wood, Gravel, and Water). The resulting 18 stimuli were repeated twice for a total 36 trials. Each trial was composed by a sequence of 6 steps. Table 9 shows, for each stimulus, the average value of the spectral centroid and of the peak level of the sequence of footsteps sounds. Each footstep sound was detected according to the technique for automatic footstep sounds extraction presented in [24]. A constant pace for all stimuli was chosen in order to avoid the bias on gender perception related to pace: as shown in previous studies [7,21,19], fast walking paces are more likely associated to female walkers at auditory level. In this way only the effect on gender perception of the features of each footstep sound per se could be assessed. The familiarization phase involved three stimuli simulating walking performed on mud by walkers having the three body sizes, wearing boots.

5.5. Results

The results for experiment 1, 2, 3, and 4 are illustrated in Tables 5–9 respectively.

As ar as the experiment 1 is concerned, the results presented in Table 5 show that 17 out of the 20 stimuli received the highest average plausibility score. Results for all stimuli belonging to each

Table 5
Results of experiment 1. Mean and standard errors of the plausibility evaluations of locomotions.

Stimulus	Response				
	W	WWS	R	J	S
Results of condition dress shoes-wood					
W	**6.62 ± 0.72**	0.39 ± 0.18	0 ± 0	1.99 ± 0.72	0 ± 0
WWS	1.96 ± 0.65	**5.49 ± 0.9**	0.24 ± 0.24	1.76 ± 0.71	0 ± 0
R	0.59 ± 0.41	0 ± 0	**6.88 ± 0.74**	1.34 ± 0.51	0 ± 0
J	4.63 ± 0.89	0 ± 0	0.76 ± 0.46	**3.72 ± 0.91**	0 ± 0
S	0.32 ± 0.24	0 ± 0	0 ± 0	0 ± 0	**8.28 ± 0.56**
Results of condition sneakers-wood					
W	**5.47 ± 0.75**	0.35 ± 0.32	0.17 ± 0.17	3.26 ± 0.85	0 ± 0
WWS	0.36 ± 0.24	**6.51 ± 0.86**	0.3 ± 0.16	1.37 ± 0.6	0 ± 0
R	0.53 ± 0.37	0 ± 0	**6.56 ± 0.79**	1.17 ± 0.57	0 ± 0
J	4.52 ± 0.9	0 ± 0	0.25 ± 0.15	**4.57 ± 0.88**	0 ± 0
S	0.07 ± 0.07	0 ± 0	0 ± 0	0 ± 0	**8.4 ± 0.49**
Results of condition dress shoes-gravel					
W	**6.96 ± 0.66**	1.36 ± 0.6	0 ± 0	1.08 ± 0.52	0 ± 0
WWS	0.69 ± 0.43	**7.13 ± 0.81**	0.1 ± 0.07	1.84 ± 0.62	0 ± 0
R	0.37 ± 0.26	0.35 ± 0.35	**6.76 ± 0.77**	1.58 ± 0.54	0.04 ± 0.04
J	5.22 ± 0.85	0.77 ± 0.47	0.34 ± 0.34	**2.51 ± 0.82**	0 ± 0
S	0.12 ± 0.12	0.09 ± 0.09	0 ± 0	0 ± 0	**8.6 ± 0.48**
Results of condition sneakers-gravel					
W	**7.82 ± 0.46**	1.07 ± 0.44	0 ± 0	0.38 ± 0.32	0.69 ± 0.3
WWS	0.77 ± 0.44	**6.64 ± 0.78**	0.1 ± 0.1	1.36 ± 0.61	0.11 ± 0.11
R	0.61 ± 0.42	0.38 ± 0.36	**6.73 ± 0.79**	1.42 ± 0.57	0 ± 0
J	6.2 ± 0.78	0.53 ± 0.36	0.69 ± 0.43	**1.66 ± 0.68**	0 ± 0
S	0.54 ± 0.3	0.17 ± 0.11	0 ± 0	0 ± 0	**8.4 ± 0.46**

W = walking, WWS = walking with scuffs, R = running, J = jumping in place, S = sliding. The results for the scale corresponding to each simulated locomotion condition are presented bold.

of the five locomotion conditions were grouped together and were then subjected to a Friedman test for each of the five locomotion scales. Post-hoc analyses were performed by using the Wilcoxon-Nemenyi-McDonald-Thompson test. A significant main effect was found for the condition walking ($\chi^2(4) = 36.24, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other locomotions (all $p < 0.001$). A significant main effect was found for the condition walking with scuffs ($\chi^2(4) = 24.1, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other locomotions (all $p < 0.001$). A significant main effect was found for the condition running ($\chi^2(4) = 30, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other locomotions (all $p < 0.001$). A significant main effect was found for the condition jumping in place ($\chi^2(4) = 34.86, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of walking with scuffs, running, and sliding (all $p < 0.001$), but significantly lower than walking ($p < 0.001$). A significant main effect was found for the condition sliding ($\chi^2(4) = 42.3, p < 0.001$); the post hoc analysis revealed that the plausibility scores for the walking condition were significantly greater than the plausibility scores of all the other locomotions (all $p < 0.001$).

Concerning experiment 2, the results presented in Tables 6 and 7 show that 29 out of the 42 stimuli received the highest average plausibility score. As far as the statistical analysis is concerned, rather than investigating the significant differences in the evaluations for each material compared on each evaluation scale, an analysis on the differences between the four typologies of surface materials involved (i.e., solid, aggregate, liquid, and hybrid) was performed. The averages presented in Tables 6 and 7 were grouped together by surface typology and were then subjected to a Kruskal Wallis test for each of the four surface typology scales. For this purpose, the surface materials that belonged to the analyzed typology and that received average evaluations equal to null were discarded. Post-hoc analyses were performed by using the Mann–Whitney tests with Bonferroni correction. A significant effect of surface typology on the plausibility evaluations was found for the condition solid ($\chi^2(3) = 146.3, p < 0.001$). The post hoc test showed that the evaluations for solid surfaces were significantly greater than those of all other typologies (all $p < 0.001$). A significant effect was found for the condition aggregate ($\chi^2(3) = 246.2, p < 0.001$). The post hoc test showed that the evaluations for aggregate surfaces were significantly greater than those of all other typologies (all $p < 0.001$). A significant effect was found for the condition liquid ($\chi^2(3) = 45.4, p < 0.001$). The post hoc test showed that the evaluations for solid surfaces were significantly greater than those of all other typologies (all $p < 0.001$). A significant effect was found for the condition ($\chi^2(3) = 83.7, p < 0.001$). The post hoc test showed that the evaluations for solid surfaces were significantly greater than those of all other typologies (all $p < 0.05$).

As ar as the experiment 3 is concerned, the results presented in Table 8 show that 27 out of the 33 stimuli received the highest average plausibility score. Results for all stimuli belonging to each of the seven shoes conditions were grouped together and were then subjected to a Friedman test for each of the seven shoes scales. Post-hoc analyses were performed by using the Wilcoxon-Nemenyi-McDonald-Thompson test. A significant main effect was found for the condition dress shoes ($\chi^2(6) = 45.5, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.05$) with the exception of boots. A significant main effect was found for the condition squeaking dress shoes ($\chi^2(6) = 35.8, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.001$). A significant main effect was found for the condition high heels

Table 6
Results of experiment 2. Mean and standard errors of the plausibility evaluations of surface materials, for stimuli involving dress shoes.

Stimulus	Response																				
	WD	CWD	CN	CNP	MB	MT	LV	DLV	SA	SDS	CS	DI	FU	WFU	CGR	WCGR	FGR	WFGR	MU	LWT	DWT
WD	**6.83 ± 0.67**	0 ± 0	3.85 ± 0.76	0 ± 0	2.62 ± 0.54	0.16 ± 0.16	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CWD	0.63 ± 0.3	**8.61 ± 0.54**	0.02 ± 0.02	0.53 ± 0.37	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CN	4.11 ± 0.74	0 ± 0	**4.71 ± 0.75**	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CNP	1.47 ± 0.5	2.86 ± 0.77	1.78 ± 0.57	**3.99 ± 0.85**	2.26 ± 0.61	0.36 ± 0.35	4.83±0.74	0.24 ± 0.12	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
MB	2.43 ± 0.68	0 ± 0	4.33 ± 0.74	0 ± 0	**6.59 ± 0.76**	1.08 ± 0.52	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
MT	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	**7.73 ± 0.6**	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
LV	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	**1.5 ± 0.57**	2.4 ± 0.63	2.61 ± 0.58	2.45 ± 0.65	3.18 ± 0.76	0.37 ± 0.15	0.87 ± 0.33	0 ± 0	0.44 ± 0.41	0.01 ± 0.01	3.63 ± 0.6	0.82 ± 0.47	0 ± 0	0 ± 0	0 ± 0
DLV	0 ± 0	0 ± 0	0 ± 0	0.53 ± 0.42	0 ± 0	0 ± 0	1.57 ± 0.58	**2.17 ± 0.55**	1.84 ± 0.65	0 ± 0	1.17 ± 0.45	0.51 ± 0.31	1.47 ± 0.55	0.91 ± 0.42	0.57 ± 0.39	0.12 ± 0.12	2.98 ± 0.74	0.92 ± 0.33	0.07 ± 0.7	0 ± 0	0 ± 0
SA	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.16 ± 0.16	0.3 ± 0.2	0.69 ± 0.25	**2.66 ± 0.69**	2.69 ± 0.79	3.56 ± 0.78	0.75 ± 0.36	0.65 ± 0.3	0 ± 0	0.22 ± 0.19	0.01 ± 0.01	2.93 ± 0.73	1.04 ± 0.52	0 ± 0	0 ± 0	0 ± 0
SDS	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.16 ± 0.1	0.11 ± 0.1	**4.79 ± 0.87**	4.2 ± 0.81	0 ± 0	0.07 ± 0.07	0.12 ± 0.12	1.37 ± 0.63	0.35 ± 0.35	0.03 ± 0.01	0 ± 0	0.18 ± 0.12	0 ± 0	0 ± 0
CS	0 ± 0	0 ± 0	0 ± 0	0.08 ± 0.06	0 ± 0	0.01 ± 0.01	0.13 ± 0.08	0.34 ± 0.16	0.18 ± 0.13	1.92 ± 0.66	**6.12 ± 0.77**	0.92 ± 0.57	1 ± 0.44	0 ± 0	2.02 ± 0.73	0.65 ± 0.4	0.12 ± 0.08	0.02 ± 0.01	0.05 ± 0.04	0 ± 0	0 ± 0
DI	0 ± 0	0 ± 0	0.01 ± 0.01	0.47 ± 0.28	0 ± 0	0 ± 0	0.42 ± 0.31	0.05 ± 0.05	0.18 ± 0.12	1.36 ± 0.58	3.39 ± 0.72	**1.74 ± 0.6**	1.3 ± 0.51	0 ± 0	2.39 ± 0.71	0.8 ± 0.41	0.46 ± 0.22	0.28 ± 0.2	0.07 ± 0.07	0 ± 0	0 ± 0
FU	0 ± 0	0.38 ± 0.38	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1.29 ± 0.54	1.41 ± 0.52	0.1 ± 0.1	1.18 ± 0.58	0.28 ± 0.7	1.08 ± 0.44	**4.43 ± 0.77**	0.07 ± 0.07	1.36 ± 0.49	0.16 ± 0.16	0.26 ± 0.18	0.35 ± 0.23	0 ± 0	0 ± 0	0 ± 0
WFU	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.3 ± 0.3	0 ± 0	0.2 ± 0.2	1.1 ± 0.47	3.47 ± 0.73	0.4 ± 0.29	**3.47 ± 0.73**	0 ± 0	2.43 ± 0.6	0 ± 0	1.44 ± 0.54	2.6 ± 0.71	1.51 ± 0.59	1.25 ± 0.59
CGR	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.32 ± 0.26	1.28 ± 0.49	0 ± 0	0.1 ± 0.1	1.7 ± 0.63	0.36 ± 0.21	1.29 ± 0.47	0 ± 0	**6.56 ± 0.74**	1.52 ± 0.49	2.33 ± 0.6	0.84 ± 0.42	0 ± 0	0 ± 0	0 ± 0
WCGR	0 ± 0	0 ± 0	0 ± 0	0.27 ± 0.18	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.15 ± 0.15	0.35 ± 0.3	1.68 ± 0.58	0 ± 0	1.68 ± 0.58	0.39 ± 0.32	**6.23 ± 0.66**	0 ± 0	1.48 ± 0.47	0.69 ± 0.38	2.5 ± 0.75	0.43 ± 0.23
FGR	0 ± 0	0 ± 0	0 ± 0	1.49 ± 0.64	0 ± 0	0 ± 0	0.15 ± 0.15	0.52 ± 0.27	0.48 ± 0.29	0 ± 0	0.27 ± 0.15	0.62 ± 0.29	0.87 ± 0.46	0.05 ± 0.05	1.52 ± 0.58	1.22 ± 0.58	**5.49 ± 0.69**	1.87 ± 0.69	0.03 ± 0.03	0.13 ± 0.13	0 ± 0
WFGR	0 ± 0	0 ± 0	0 ± 0	0.68 ± 0.42	0 ± 0	0 ± 0	0.12 ± 0.12	0.1 ± 0.1	0 ± 0	0 ± 0	0.17 ± 0.17	0 ± 0	0.15 ± 0.15	0.98 ± 0.51	0 ± 0	0.88 ± 0.43	0.36 ± 0.36	**5.49 ± 0.78**	0.91 ± 0.37	1.93 ± 0.58	0.77 ± 0.53
MU	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.2 ± 0.14	0 ± 0	0 ± 0	1.59 ± 0.64	1.08 ± 0.42	0 ± 0	0.08 ± 0.08	1.33 ± 0.43	0.05 ± 0.05	2.26 ± 0.68	0.01 ± 0.01	1.77 ± 0.52	**2.08 ± 0.64**	2.06 ± 0.66	0.02 ± 0.02
LWT	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.07 ± 0.07	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.1 ± 0.1	**5.98 ± 0.79**	1.73 ± 0.59
DWT	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.06 ± 0.06	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.17 ± 0.14	1.38 ± 0.46	**7.03 ± 0.77**

WD = wood, CWD = creaking wood, CN = concrete, CNP = concrete plus pebbles, MB = marble, MT = metal, LV = leaves, DLV = dry leaves, SA = sand, SDS = soft deep snow, CS = crunchy snow, DI = dirt, FU = forest underbrush, WFU = wet forest underbrush, CGR = coarse gravel, WCGR = wet coarse gravel, FGR = fine gravel, WFGR = wet fine gravel, MU = mud, LWT = low water puddle, DWT = deep water puddle. The results for the scale corresponding to each simulated surface material condition are presented bold.

Table 7
Results of experiment 2. Mean and standard errors of the plausibility evaluations of surface materials, for stimuli involving sneakers.

Stimulus	Response																				
	WD	CWD	CN	CNP	MB	MT	LV	DLV	SA	SDS	CS	DI	FU	WFU	CGR	WCGR	FGR	WFGR	MU	LWT	DWT
WD	**2.43 ± 0.63**	0.04 ± 0.04	2.56 ± 0.41	0.84 ± 0.38	0.9 ± 0.46	0.04 ± 0.04	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.82 ± 0.51	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CWD	1.24 ± 0.49	**8.52 ± 0.58**	0 ± 0	0.18 ± 0.18	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CN	1.51 ± 0.44	0.05 ± 0.05	**2.1 ± 0.49**	2.56 ± 0.61	0.25 ± 0.16	0.01 ± 0.01	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.25 ± 0.23	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CNP	0.27 ± 0.16	0.39 ± 0.18	0.77 ± 0.32	**6.48 ± 0.46**	0.1 ± 0.06	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.48 ± 0.34	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
MB	1.26 ± 0.34	0.2 ± 0.18	1.92 ± 0.6	2.34 ± 0.44	**0.83 ± 0.34**	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.29 ± 0.29	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
MT	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.19 ± 0.16	**8.24 ± 0.48**	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
LV	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	**0.79 ± 0.39**	1.66 ± 0.52	6.03 ± 0.68	2.42 ± 0.65	1.45 ± 0.4	0.41 ± 0.25	0.69 ± 0.45	0.41 ± 0.41	0 ± 0	0 ± 0	2.1 ± 0.6	0.48 ± 0.38	0 ± 0	0 ± 0	0 ± 0
DLV	0 ± 0	0 ± 0	0 ± 0	0.34 ± 0.34	0 ± 0	0 ± 0	1.46 ± 0.51	**3.9 ± 0.75**	1.52 ± 0.53	0.35 ± 0.34	0.85 ± 0.42	0.71 ± 0.36	1.67 ± 0.45	0.33 ± 0.31	0.11 ± 0.08	0 ± 0	2.31 ± 0.64	0.8 ± 0.49	0 ± 0	0 ± 0	0 ± 0
SA	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.44 ± 0.36	0.46 ± 0.28	**3.35 ± 0.65**	3.34 ± 0.76	3.52 ± 0.8	0.67 ± 0.38	0.49 ± 0.29	0 ± 0	0.02 ± 0.02	0 ± 0	2.75 ± 0.72	0.1 ± 0.07	0 ± 0	0 ± 0	0 ± 0
SDS	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.17 ± 0.12	**6.11 ± 0.86**	3.7 ± 0.86	0.79 ± 0.42	0.4 ± 0.23	0.83 ± 0.48	0 ± 0	0 ± 0	0.13 ± 0.13	0.38 ± 0.38	0 ± 0	0 ± 0	0 ± 0
CS	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.65 ± 0.43	0.2 ± 0.12	0.3 ± 0.18	3.65 ± 0.7	**5.23 ± 0.73**	0.73 ± 0.41	0.42 ± 0.23	0.67 ± 0.48	0.91 ± 0.51	0.26 ± 0.26	0.2 ± 0.16	0 ± 0	0.09 ± 0.09	0 ± 0	0 ± 0
DI	0.21 ± 0.21	0 ± 0	0 ± 0	0.12 ± 0.09	0 ± 0	0 ± 0	0.49 ± 0.35	0.34 ± 0.21	0.09 ± 0.09	1.99 ± 0.53	3.44 ± 0.73	**2.04 ± 0.65**	0.45 ± 0.4	0.7 ± 0.39	0.11 ± 0.09	0.48 ± 0.3	0.07 ± 0.06	0.5 ± 0.33	0 ± 0	0 ± 0	0 ± 0
FU	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	1.53 ± 0.52	0.94 ± 0.4	0.11 ± 0.07	0.9 ± 0.49	1.59 ± 0.47	1.74 ± 0.53	**6.53 ± 0.62**	0.49 ± 0.31	0.67 ± 0.38	0.04 ± 0.04	1.21 ± 0.51	0 ± 0	0 ± 0	0 ± 0	0 ± 0
WFU	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.02 ± 0.02	0.48 ± 0.39	0 ± 0	0.13 ± 0.1	0.48 ± 0.33	0 ± 0	0.03 ± 0.03	**3.52 ± 0.81**	0 ± 0	1.55 ± 0.59	0 ± 0	1.87 ± 0.52	1.61 ± 0.5	2.52 ± 0.6	0.03 ± 0.03
CGR	0 ± 0	0 ± 0	0 ± 0	0.3 ± 0.28	0 ± 0	0 ± 0	0.3 ± 0.2	1.41 ± 0.58	0.14 ± 0.11	0.12 ± 0.09	1.22 ± 0.41	1.2 ± 0.54	3.3 ± 0.73	0 ± 0	**4.55 ± 0.8**	0.79 ± 0.44	2.23 ± 0.61	1 ± 0.56	0.03 ± 0.03	0 ± 0	0 ± 0
WCGR	0 ± 0	0 ± 0	0 ± 0	0.48 ± 0.32	0 ± 0	0 ± 0	0.59 ± 0.36	0 ± 0	0.02 ± 0.02	0.23 ± 0.22	0.21 ± 0.18	0.12 ± 0.12	0.23 ± 0.16	1.89 ± 0.64	1.38 ± 0.57	**4.11 ± 0.74**	0.66 ± 0.36	3.41 ± 0.73	1.04 ± 0.46	0.93 ± 0.42	0.31 ± 0.26
FGR	0 ± 0	0 ± 0	0 ± 0	0.66 ± 0.3	0.16 ± 0.16	0 ± 0	0.15 ± 0.11	0.51 ± 0.33	0.32 ± 0.3	0 ± 0	0.49 ± 0.24	0.83 ± 0.39	0.97 ± 0.42	0.25 ± 0.25	1.36 ± 0.56	0.76 ± 0.46	**7.04 ± 0.61**	1.68 ± 0.64	0.17 ± 0.17	0.13 ± 0.13	0 ± 0
WFGR	0 ± 0	0 ± 0	0 ± 0	0.39 ± 0.29	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.13 ± 0.13	0 ± 0	0 ± 0	0.86 ± 0.33	0.32 ± 0.32	1.6 ± 0.55	0.69 ± 0.49	**7.02 ± 0.6**	1.05 ± 0.49	2.64 ± 0.68	0.63 ± 0.34
MU	0 ± 0	0 ± 0	0.02 ± 0.02	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.13 ± 0.13	1.7 ± 0.6	1.27 ± 0.45	0.45 ± 0.25	0.11 ± 0.06	1.14 ± 0.35	0.13 ± 0.13	0.57 ± 0.4	0.46 ± 0.35	1.71 ± 0.56	**2.56 ± 0.77**	0.94 ± 0.45	0 ± 0
LWT	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.05 ± 0.05	0 ± 0	0.2 ± 0.2	0 ± 0	0 ± 0	0 ± 0	0.01 ± 0.01	0 ± 0	0.2 ± 0.18	1.86 ± 0.64	**6.21 ± 0.82**	0.24 ± 0.2
DWT	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0.16 ± 0.11	0.21 ± 0.17	1.85 ± 0.56	**6.77 ± 0.69**

WD = wood, CWD = creaking wood, CN = concrete, CNP = concrete plus pebbles, MB = marble, MT = metal, LV = leaves, DLV = dry leaves, SA = sand, SDS = soft deep snow, CS = crunchy snow, DI = dirt, FU = forest underbrush, WFU = wet forest underbrush, CGR = coarse gravel, WCGR = wet coarse gravel, FGR = fine gravel, WFGR = wet fine gravel, MU = mud, LWT = low water puddle, DWT = deep water puddle. The results for the scale corresponding to each simulated surface material condition are presented bold.

Table 8
Results of experiment 3. Mean and standard errors of the plausibility evaluations of shoe types.

Stimulus	Response						
	DS	SDS	HH	B	BB	SN	SSN
Results of condition wood-male walker							
DS	**4.64 ± 0.8**	0.73 ± 0.47	4.32 ± 0.89	2.46 ± 0.65	0.01 ± 0.01	0.05 ± 0.05	0 ± 0
SDS	1.16 ± 0.5	**7.56 ± 0.67**	0.86 ± 0.47	0.56 ± 0.39	0.32 ± 0.3	0 ± 0	0.95 ± 0.45
B	1.12 ± 0.35	0.39 ± 0.23	0 ± 0	**6 ± 0.6**	0.55 ± 0.4	0.37 ± 0.27	0 ± 0
BB	1 ± 0.52	0.73 ± 0.43	1.15 ± 0.44	0.04 ± 0.03	**8.16 ± 0.55**	0 ± 0	0 ± 0
SN	1.56 ± 0.48	0.36 ± 0.23	0 ± 0	3.47 ± 0.73	0 ± 0	**3.13 ± 0.65**	0 ± 0
SSN	0 ± 0	2.11 ± 0.54	0 ± 0	0.07 ± 0.05	0 ± 0	0.36 ± 0.25	**6.6 ± 0.54**
Results of condition wood-female walker							
HH	3.06 ± 0.72	0.6 ± 0.42	**8.08 ± 0.43**	0.22 ± 0.1	0 ± 0	0 ± 0	0 ± 0
B	0.97 ± 0.29	0.35 ± 0.24	0.2 ± 0.14	**5.49 ± 0.63**	0.94 ± 0.54	0.29 ± 0.22	0 ± 0
BB	0.58 ± 0.4	0.22 ± 0.22	1.37 ± 0.54	0.08 ± 0.08	**7.63 ± 0.65**	0 ± 0	0 ± 0
SN	1.48 ± 0.46	0.1 ± 0.07	0.45 ± 0.26	1.9 ± 0.57	0 ± 0	**3.61 ± 0.67**	0.2 ± 0.14
SSN	0 ± 0	1.82 ± 0.49	0 ± 0	0.07 ± 0.07	0 ± 0	0.31 ± 0.22	**6.3 ± 0.59**
Results of condition gravel-male walker							
DS	**3.41 ± 0.63**	0.55 ± 0.4	2.24 ± 0.66	2.89 ± 0.71	0.75 ± 0.4	0.94 ± 0.35	0.68 ± 0.38
SDS	0.95 ± 0.5	**5.23 ± 0.8**	0.6.±0.36	1.1 ± 0.41	0.68 ± 0.43	1.2 ± 0.5	3.09 ± 0.77
B	2.8 ± 0.7	1.45 ± 0.63	1.78 ± 0.71	**4.16 ± 0.84**	1.01 ± 0.52	1.43 ± 0.42	0.35 ± 0.25
BB	0.52 ± 0.38	0.35 ± 0.32	0.61 ± 0.39	0.74 ± 0.48	**7.53 ± 0.70**	0.35 ± 0.22	0.47 ± 0.32
SN	3.1 ± 0.73	1.15 ± 0.55	1.45 ± 0.59	3.23 ± 0.8	0.69 ± 0.47	**3.55 ± 0.72**	0.9 ± 0.42
SSN	0.66 ± 0.27	4.08 ± 0.66	0.01 ± 0.01	1.14 ± 0.51	0.05 ± 0.05	1.64 ± 0.55	**4.73 ± 0.66**
Results of condition gravel-female walker							
HH	2.18 ± 0.67	1.15 ± 0.5	**0.61 ± 0.29**	1.57 ± 0.57	0.27 ± 0.2	4.23 ± 0.73	1.39 ± 0.52
B	4.24 ± 0.69	0.71 ± 0.4	1.53 ± 0.61	**3.31 ± 0.76**	0.82 ± 0.53	2.38 ± 0.57	1.04 ± 0.46
BB	0.15 ± 0.09	0.08 ± 0.08	0.19 ± 0.11	0.3 ± 0.23	**7.1 ± 0.74**	0.16 ± 0.1	0.43 ± 0.3
SN	2.77 ± 0.71	0.83 ± 0.47	1.45 ± 0.5	1.4 ± 0.57	0 ± 0	**4.29 ± 0.63**	1.9 ± 0.59
SSN	0.23 ± 0.22	3.08 ± 0.7	0.35 ± 0.24	0.59 ± 0.41	0.01 ± 0.01	1.86 ± 0.61	**6.38 ± 0.62**
Results of condition water-male walker							
DS	**3.15 ± 0.64**	0.95 ± 0.46	1.88 ± 0.69	2.12 ± 0.64	0.55 ± 0.43	1.16 ± 0.39	0.89 ± 0.43
SDS	0.94 ± 0.46	**2.55 ± 0.58**	0.94 ± 0.51	1.31 ± 0.44	0.77 ± 0.46	0.06 ± 0.05	3.54 ± 0.8
B	2.06 ± 0.59	1.42 ± 0.54	0.9 ± 0.45	**2.66 ± 0.61**	0.2 ± 0.12	1.8 ± 0.61	1.5 ± 0.6
BB	0.5 ± 0.2	0.35 ± 0.22	0.16 ± 0.13	0.45 ± 0.28	**6.48 ± 0.85**	0.31 ± 0.31	0.17 ± 0.17
SN	1.69 ± 0.61	1.04 ± 0.5	0.63 ± 0.37	1.68 ± 0.48	0.29 ± 0.19	**3.11 ± 0.7**	1.94 ± 0.62
SSN	0.77 ± 0.44	1.89 ± 0.53	0.85 ± 0.52	0.6 ± 0.33	0.59 ± 0.34	1.17 ± 0.48	**5.36 ± 0.74**
Results of condition water-female walker							
HH	1.68 ± 0.59	0.6 ± 0.41	**4.2 ± 0.71**	1.24 ± 0.48	0.41 ± 0.31	0.97 ± 0.38	0 ± 0
B	2.33 ± 0.63	2.12 ± 0.66	2.14 ± 0.61	**1.32 ± 0.5**	0.84 ± 0.38	1.67 ± 0.42	0.83 ± 0.37
BB	0.94 ± 0.46	0.55 ± 0.38	1.11 ± 0.56	0.63 ± 0.25	**5.81 ± 0.87**	0.37 ± 0.21	0 ± 0
SN	0.82 ± 0.35	1.43 ± 0.63	0.31 ± 0.19	1.65 ± 0.49	0.29 ± 0.21	**3.93 ± 0.6**	1.77 ± 0.57
SSN	0.35 ± 0.24	2.31 ± 0.64	0.4 ± 0.3	0.73 ± 0.42	0.06 ± 0.06	1.64 ± 0.6	**4.67 ± 0.69**

DS = dress shoes, SDS = squeaking dress shoes, HH = high heels, B = boots, BB = boots with buckles, SSN = squeaking sneakers, SN = sneakers. The results for the scale corresponding to each simulated shoe type condition are presented bold.

($\chi^2(6) = 41.5, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.001$) with the exception of dress shoes. A significant main effect was found for the condition boots ($\chi^2(6) = 36.24, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.001$). A significant main effect was found for the condition boots with buckles ($\chi^2(6) = 32.2, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.001$). A significant main effect was found for the condition sneakers ($\chi^2(6) = 46.4, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.001$). A significant main effect was found for the condition squeaking sneakers ($\chi^2(6) = 50, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of all the other shoes (all $p < 0.001$).

Concerning experiment 4, the results presented in Table 9 show that 14 out of the 18 stimuli received the highest average plausibility score. Results for all stimuli belonging to each of the three body size conditions were grouped together and were then subjected to a Friedman test for each of the three body size scales. Post-hoc analyses were performed by using the Wilcoxon-Nemenyi-McDonald-Thompson test. A significant main effect was found for the condition big ($\chi^2(2) = 15.5, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of the small body size ($p < 0.001$). A significant main effect was found for the condition medium ($\chi^2(2) = 18.1, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of the other two body sizes (both $p < 0.001$). A significant main effect was found for the condition small ($\chi^2(2) = 22.1, p < 0.001$); the post hoc analysis revealed that the plausibility scores for such condition were significantly greater than the plausibility scores of the other two body sizes (both $p < 0.001$).

A linear mixed-effects model analysis was performed, separately for each of the three surface materials involved (Wood, Gravel, and Water), in order to search for correlations between the plausibility evaluations of femaleness and the average spectral centroid and peak level of the footstep sounds contained in each stimulus (see Table 9), as well as between femaleness judgments and the body sizes simulated in each stimulus. For the latter

Table 9
Results of experiment 4. Mean and standard errors of the plausibility evaluations of body size and gender.

Stimulus	Centroid (Hz)	Peak (dB)	Response			
			B	M	S	F
Results of condition dress shoes-wood						
B	379.9	−5.951	**6 ± 0.68**	5 ± 0.62	0.52 ± 0.24	2.66 ± 0.61
M	644	−15.237	3.13 ± 0.72	**6.92 ± 055**	1.33 ± 0.36	4.77 ± 0.69
S	1472.1	−26.736	0.86 ± 0.46	4.18 ± 0.63	**6.63 ± 0.74**	7.08 ± 0.68
Results of condition dress shoes-gravel						
B	9797.2	−20.333	**4.19 ± 0.7**	7.44 ± 0.47	0.6 ± 0.21	3.69 ± 0.35
M	9944.6	−18.819	3.32 ± 0.71	**6.02 ± 0.59**	1.51 ± 0.58	3.93 ± 0.49
S	10,456	−26.961	0.56 ± 0.28	2.51 ± 0.59	**6.74 ± 0.7**	6.39 ± 0.43
Results of condition dress shoes-water						
B	888.1	−13.576	**4.21 ± 0.71**	4.66 ± 0.63	2.1 ± 0.78	3.62 ± 0.54
M	1215.4	−13.308	0.94 ± 0.33	**5.71 ± 0.64**	4.68 ± 0.81	5.82 ± 0.48
S	2327.7	−28.832	0.62 ± 0.36	3.23 ± 0.54	**7.29 ± 0.57**	7.42 ± 0.48
Results of condition sneakers-wood						
B	939.8	−19.392	**6.12 ± 0.63**	4.19 ± 0.64	0.73 ± 0.27	1.6 ± 0.3
M	1715.2	−27.497	4.18 ± 0.81	**5.06 ± 0.75**	1.96 ± 0.71	2.63 ± 0.56
S	3262.4	−39.601	0.8 ± 0.43	2.83 ± 0.6	**7.99 ± 0.67**	6.03 ± 0.69
Results of condition sneakers-gravel						
B	9781.8	−17.815	**4.97 ± 0.78**	5.16 ± 0.55	1.26 ± 0.52	3.27 ± 0.36
M	9961.2	−17.122	2.55 ± 0.65	**6.15 ± 0.64**	1.95 ± 0.59	4.78 ± 0.44
S	10,493	−24.787	0.52 ± 0.21	2.99 ± 0.67	**6.7 ± 0.74**	7.46 ± 0.53
Results of condition sneakers-water						
B	1271	−18.633	**1.76 ± 0.45**	6.87 ± 0.64	2.45 ± 0.68	4.28 ± 0.61
M	1392.6	−22.392	1.02 ± 0.4	**5.66 ± 0.72**	4.95 ± 0.83	6.19 ± 0.49
S	2434.8	−32.385	0.11 ± 0.07	2.21 ± 0.52	**7.94 ± 0.57**	7.76 ± 0.46

B = big, M = medium, S = small, F = femaleness. For each stimulus the average value of the spectral centroid and of the peak level of the sequence of footsteps sounds are presented. The results for the scale corresponding to each simulated body size condition are presented bold.

Table 10
Significant correlations in the linear mixed-effect model analysis. All t-tests are significant at $p < 0.001$.

Predicted by predictor	β weight	t-test
Results of condition wood		
Femaleness by spectral centroid	0.001	$t(131) = 3.68$
Femaleness by peak level	−0.096	$t(131) = 3.96$
Femaleness by body size	−2.214	$t(131) = -8.09$
Results of condition gravel		
Femaleness by spectral centroid	0.005	$t(131) = 8.57$
Femaleness by peak level	−0.338	$t(131) = -6.64$
Femaleness by body size	−1.724	$t(131) = -7.95$
Results of condition water		
Femaleness by spectral centroid	0.002	$t(131) = 7.461$
Femaleness by peak level	−0.176	$t(131) = -6.74$
Femaleness by body size	−1.820	$t(131) = -8.41$

regression, the three body sizes big, medium, and small, were coded 3, 2, and 1 respectively. Such analyses revealed that the femaleness scores were linearly related to the sounds' spectral centroid and peak level, and simulated body size. Table 10 reports details of the significant linear correlations.

5.6. Discussion

The results of experiment 1 show that all the simulated locomotions were correctly identified and well discriminated between each other, with the exception of jumping in place, which was mainly interpreted as walking, especially in presence of the gravel surface material. This results, however, is not surprising since one of the main discriminant between walking and jumping in place is the variation of the distance from the listener of the moving sound source in the walking case, which was not modeled by the synthesizer.

The results of experiment 2 show that the majority of the simulated surface materials were correctly identified. As expected, better identifications for solid surfaces were found in presence of dress shoes compared to sneakers. The statistical analysis shows that the four typologies of surface materials (solid, aggregate, liquid, and hybrid) were very well discriminated. This result perfectly parallels that of previous studies performed involving real stimuli listened by participants walking on real surface materials, recordings of real walks, as well as synthesized stimuli [27,28].

The results of experiment 3 show that all the simulated shoe types were correctly identified and well discriminated between each other with the exception of dress shoes (which was not significanlty distinguished from boots) and of high heels (which was not significanlty distinguished from dress shoes).

The results of experiment 4 show that the three simulated body sizes were correctly identified and well discriminated between each other, with the exception of the big one, which was not significanlty distinguished from the medium one. Femaleness perception was modulated by the simulated body size as well as by the peak level and the spectral content of the sounds. The latter result confirms the finding of previous studies on gender identification from real and synthesized walking sounds [7,21,19]. It is worth to notice that all the stimuli involved in this experiment were synthesized with the same temporal distance between steps: the femaleness perception could have been increased/decreased by varying also such a synthesis parameter. Nevertheless, the choice of not acting on this parameter was due to the goal of assessing the effect on gender perception of the spectral content and temporal evolution of the involved footsteps sounds.

6. General discussion and conclusion

In this paper a series of algorithmic solutions for the synthesis of footstep sounds have been proposed. The synthesis was achieved by means of the control of physical and physically inspired models in order to generate different types of foot–floor interactions, various kinds of shoes, and ground materials. In addition, some anthropomorphic features of the walker were simulated. A great variety of footstep sounds can be created by combining all these factors.

The synthesis was conceived in order to accomplish simulations that are valid from the ecological point of view and are perceptually plausible. For this purpose, an approach based on separate rendering, and subsequent combination, of structural and transformational invariants was adopted. Moreover, the analysis-by-synthesis approach was followed. The whole design was driven also by the goal of achieving a real-time parametric control. To this end, the cartoonification technique was adopted, which made the simulations computational efficient. Therefore, the developed synthesizer is suitable for interactive applications involving physical locomotion where timeliness is a fundamental requisite [78]. It can be utilized in conjunction with locomotion interfaces, such as instrumented shoes [75] or augmented floors [77,18], that are capable of displaying not only auditory but also plantar vibrotactile feedback. These interfaces leverage the control of the vibrotactile feedback by means of techniques of the sound generation. Such an approach, has been proved to be successful in previous research [79,56,80]. It is motivated by the fact that in real life the mechanical source of vibration is the same for both the auditory and tactile modality. What differs are the transmission medium and the organs used to pick up the vibrations: air and ears for audition, shoes and foot mechanoreceptors for touch.

In a different vein, this paper faced the problem of how to control the presented synthesis algorithms. Firstly, a control method was proposed in order to generate sequences of footstep sounds. It was inspired by biomechanical models of locomotions and results of footsteps sounds perception research. It represents an improvement of the state-of-the art of existing footstep sounds sequencers [6,7,29].

Secondly, the design choices underlying the tuning of the synthesis parameters were illustrated. On the one hand, they were driven by results of sound perception research. On the other hand, by the author's design choices, which were anyhow based on common everyday experience. Aesthetics is a topic that is receiving growing attention in sound design research and aesthetic choices have been encouraged by different authors in sound design practice [81,82]. The overall synthesis approach, including the tuning of the synthesis parameters, was based on the objective to achieve perceptually compelling footstep sounds simulations (see Section 2.2), having as a reference the work of foley artists who invent, with their creativity and aesthetic choices, methods to produce plausible sounds [2].

Thirdly, a control strategy based on a three-layers hierarchy was proposed to provide users with the possibility of creating the footstep sounds they had in their mind (see Fig. 4). In the first layer, users were provided with an intuitive high-level control of the synthesis algorithms based on verbal descriptions of the sound source (i.e., type of ground material, shoe, locomotion, and person performing it, see Table 4). In the second and third layers, expert users could combine the sound models and manipulate their synthesis parameters in order to generate novel sounds not present among those currently available.

Fourthly, Section 4.5 showed control techniques for both the interactive case, where different types of locomotion interfaces are utilized, and the non-interactive case, where locomotion is passively simulated through the locomotion model presented in Section 4.1. The developed synthesizer and its control techniques can find application in several interactive and non-interactive contexts such as virtual reality (e.g., navigation in virtual environments) [83], entertainment (e.g., video games, movies) [31,79], or perceptual studies aiming at investigating the role of the action-perception loop in locomotion [80,84,85].

As far as the evaluation is concerned, firstly results showed that in each experiment the majority of stimuli received the highest plausibility score or the second highest one. Secondly, the statistical analysis showed that on average the stimuli were correctly identified and discriminated between each other. Taken together, the results of the four evaluation experiments showed that the stimuli synthesized according to the techniques and design choices presented in Sections 2, 3, and 4.5 were successful in conveying at auditory level the information regarding the types of foot–floor interaction, locomotion, surface material, shoe, as well as walker's body size and gender. The four experiments allowed to highlight which stimuli were less identified and discriminated from the others, therefore pinpointing directions for future improvement of the synthesis algorithms.

Future research will focus on the evaluation in an interactive context involving a locomotion interfaces, such as instrumented shoes or augmented floors [18] capable of accomplishing an interactive sonification [86] of a user's foot–floor interactions. By conducting an evaluation in a passive context, however, is a condition whose results that can be considered as a lower bound in the participants' performances for two reasons. Firstly, introducing interactivity in a surface material identification experiment involving a previous version of the synthesizer [10] has been demonstrated to lead to better performances compared to those achieved during a listening test [28]. Secondly, the sounds were presented devoid of a context. Despite footstep sounds can be assumed to have inherent meaning that is learnt from our everyday activities, hearing such sounds in isolation without context lacks of ecological validity and can even be confusing. For instance, the sound of a single isolated female footstep on wood using high heels could be identified as a book being dropped on a table. In this regard, the influence of contextual information provided as soundscape was assessed to improve surface material identification performances compared to when sounds were presented in isolation [32]. Similarly, other sources of contextual information are expected to play a relevant role in the identification of properties of footstep sounds, as well as on their perceived plausibility, such as vision, touch, and proprioception [80].

There are several possible extensions of the synthesizer. Thanks to the intuitive control of the locomotion model (see Section 4.1), it is possible to act on the two sound descriptors "time between steps" and "steps amplitude" (see Fig. 4) to create emotional expressive walking styles (e.g., sad, happy, etc.). Indeed, research on auditory perception of foot–floor interactions has shown that footstep sounds can convey the walker's emotional state [22,87]. Specifically, it has been shown that when walking with different emotional intentions humans make variations of timing and sound level in the same way as found in expressive music performance [88]. For example, it has been found that music performances communicating happiness and happy walking styles are characterized by a faster tempo/pace and louder sound level relative to a neutral style, while performances and walking patterns communicating sadness are characterized by slower tempo/pace and softer sound level. In addition to this, since the two mentioned parameters are controllable in real-time, the locomotion model is malleable to the emotional influence of the user manipulating the interface (similarly to the preliminary model proposed by De Witt and Bresin [7]).

Along the same line, it is possible to manipulate the sound descriptors "time between steps" to simulate different surface profile, such as bumps, holes, and flat surfaces, by leveraging the findings reported by Turchet and Serafin [31] that showed that variations in T are sufficient for this purpose. The simulation of climbing and descending staircases can be achieved by tuning several aspects of the synthesizer, such as different types of exciter to model the type of foot–floor interaction, different set of parameters controlling the ground models, variations in temporal distance between steps. Such information can be extracted by recordings of real steps occurring in those scenarios. Moreover, by leveraging the results reported by Pastore et al. [25], it might be possible even to simulate two postures of the walker, i.e. upright and stooped.

Currently, the synthesizer does not model the distance of the sound source from the listener. This is an aspect that confers realism to the proposed simulations in both interactive and non interactive contexts. In the interactive case, directionality and distance of the sounds resulting from the foot–floor interaction are naturally modeled if locomotion interfaces are used such as shoes with loudspeakers mounted on the top (e.g., [75]) or augmented floors having loudspeakers mounted under each tile (e.g., [77]). When headphones are involved, then binaural techniques might be utilized [89]. These are based on the precise rendering of the auditory cues arriving to each ear that can lead to accurate source localization. To the best author's knowledge, to date no binaural technique has been specifically defined for the rendering of footstep sounds in interactive contexts. As far as the non interactive case is concerned, when a multichannel loudspeaker system is involved the distance from the user and the simulated virtual walker might be rendered by sophisticated spatial rendering algorithms, such as ambisonics [90], capable of simulating virtual sound sources moving along tridimensional trajectories [91]. When headphones are utilized, binaural techniques can be used for the same purpose.

Furthermore, a model for multiple walkers can be implemented starting from the modeling of the locomotion of a single walker presented here. For instance, this could be achieved by exploiting the control system for clapping hands proposed in [42] which can mimick both the asynchronous and synchronized applause of a group of clappers. In principle it is possible to introduce in the model, a control of the following parameters: number of walkers, variability in the walkers' shoe types and anthropomorphic features, footstep synchronicity. In this way it will be possible to render several walking-based acoustic scenarios, ranging from soldiers' synchronous marching to people randomly walking or running in whatever environment. Finally, the synthesizer can be extended by introducing the simulation of the sounds of the friction of the walker's clothes (e.g., the rubbing of the pants) which can also be contained in acoustic locomotion signatures.

All these possible extensions will be a matter for future research. More importantly, the author looks forward to integrating any new findings from footstep sound perception research into future development of the footstep synthesizer.

Acknowledgment

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.

References

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

[2] Yewdall D. Foley: the art of footsteps, props, and cloth movement. In: Practical art of motion picture sound. Focal Press; 2003. p. 294–328.

[3] Fontana F, Visell Y, editors. Walking with the senses. Perceptual techniques for walking in simulated environments. Logos-Verlag; 2012.

[4] Cook P. Modeling bill's gait: analysis and parametric synthesis of walking sounds. In: Proceedings of the AES 22nd international conference on virtual, synthetic, and entertainment audio; 2002. p. 73–8.

[5] Cook P. Physically informed sonic modeling (phism): synthesis of percussive sounds. Comput Music J 1997;21(3):38–49.

[6] Fontana F, Bresin R. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In: Proc. colloquium on musical informatics; 2003. p. 109–114.

[7] DeWitt A, Bresin R. Sound design for affective interaction. In: Paiva A, Prada R, Picard RE, editors. Affective computing and intelligent interaction. Lecture notes in computer science, vol. 4738. Berlin, Heidelberg: Springer; 2007. p. 523–33.

[8] Farnell A. Marching onwards: procedural synthetic footsteps for video games and animation. In: Proceedings of the pure data convention; 2007.

[9] Galbraith F, Barton M. Ground loading from footsteps. J Acoust Soc Am 1970;48:1288–92.

[10] 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.

[11] 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.

[12] Gaver W. What in the world do we hear?: an ecological approach to auditory event perception. Ecol Psychol 1993;5(1):1–29.

[13] Gaver W. How do we hear in the world? Explorations in ecological acoustics. Ecol Psychol 1993;5(4):285–313.

[14] McAdams S, Bigand E. Thinking in sound: the cognitive psychology of human audition. Oxford University Press; 1993.

[15] Carello C, Wagman J, Turvey M. Acoustic specification of object properties. In: Moving image theory: ecological considerations. Carbondale: Southern Illinois University Press; 2003.

[16] Bever T, Poeppel D. Analysis by synthesis: a (re-)emerging program of research for language and vision. Biolinguistics 2010;4(2–3):174–200.

[17] Friberg A, Bresin R, Sundberg J. Overview of the kth rule system for musical performance. Adv Cognit Psychol 2006;2(2-3):145–61.

[18] Turchet L. Custom made wireless systems for footstep sounds synthesis. Appl Acoust 2014;83:22–31.

[19] Li XF, Logan RJ, Pastore RE. Perception of acoustic source characteristics: walking sounds. J Acoust Soc Am 1991;90(6):3036–49.

[20] Coward S, Stevens C. Extracting meaning from sound: nomic mappings, everyday listening, and perceiving object size from frequency. Psychol Rec 2004;54:349–64.

[21] Visell Y, Fontana F, Giordano B, Nordahl R, Serafin S, Bresin R. Sound design and perception in walking interactions. Int J Human–Comput Stud 2009;67(11):947–59.

[22] Giordano B, Bresin R. Walking and playing: what's the origin of emotional expressiveness in music? In: Proceedings of the 9th international conference on music perception & cognition; 2006. p. 436.

[23] Bresin R, DeWitt A, Papetti S, Civolani M, Fontana F. Expressive sonification of footstep sounds. In: Proc. of the interactive sonification workshop ISon 2010; 2010. p. 51–4.

[24] Turchet L, Bresin R. Effects of interactive sonification on emotionally expressive walking styles. IEEE Trans Affect Comput 2015;6(2):152–64.

[25] Pastore R, Flint J, Gaston J, Solomon M. Auditory event perception: the source–perception loop for posture in human gait. Attent Percept Psychophys 2008;70(1):13–29.

[26] Mäkelä K, Hakulinen J, Turunen M. The use of walking sounds in supporting awareness. In: Proceedings of ICAD; 2003. p. 144–7.

[27] Giordano B, Visell Y, Yao H-Y, Hayward V, Cooperstock J, McAdams S. Identification of walked-upon materials in auditory, kinesthetic, haptic and audio-haptic conditions. J Acoust Soc Am 2012;131:4002–12.

[28] Nordahl R, Serafin S, Turchet L. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proc IEEE virtual reality conference. IEEE Press; 2010. p. 147–53.

[29] Bresin R, Dahl S. Experiments on gestures: walking, running, and hitting. In: Rocchesso D, Fontana F, editors. The sounding object, Edizioni mondo estremo; 2003. p. 111–36.

[30] Friberg A. Generative rules for music performance: a formal description of a rule system. Comput Music J 1991;15(2):56–71.

[31] Turchet L, Serafin S. Temporal and amplitude aspects in sonically simulating walking over bumps, holes and flat surfaces. Acoust Sci Technol 2012;33(5):1–10.

[32] 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.

[33] Lutfi R. Human sound source identification. In: Yost W, Popper A, Fay R, editors. Auditory perception of sound sources. Springer handbook of auditory research, vol. 29. Springer; 2008. p. 13–42.

[34] Ballas J. Common factors in the identification of an assortment of brief everyday sounds. J Exp Psychol: Human Percept Perform 1993;19(2):250–67.

[35] Lemaitre G, Houix O, Misdariis N, Susini P. Listener expertise and sound identification influence the categorization of environmental sounds. J Exp Psychol: Appl 2010;16(1):16–32.

[36] Sonnenschein D. Sound design: the expressive power of music, voice, and sound effects in cinema. Michael Wiese Productions; 2001.

[37] Turchet L, Serafin S. Investigating the amplitude of interactive footstep sounds and soundscape reproduction. Appl Acoust 2013;74(4):566–74.

[38] Rocchesso D, Fontana F, editors. The sounding object, Edizioni mondo estremo; 2003.

[39] Giordano B, Rocchesso D, McAdams S. Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. J Exp Psychol: Human Percept Perform 2010;36(2):462–76.

[40] Giordano B, Avanzini F. Perception and synthesis of sound-generating materials. In: Multisensory softness. Springer; 2014. p. 49–84.

[41] Rocchesso D, Bresin R, Fernstrom M. Sounding objects. IEEE Multimedia 2003;10(2):42–52.

[42] Peltola L, Erkut C, Cook P, Valimaki V. Synthesis of hand clapping sounds. IEEE Trans Audio Speech Language Process 2007;15(3):1021–9.

[43] Nilsson J, Thorstensson A. Ground reaction forces at different speeds of human walking and running. Acta Physiol Scand 1989;136(2):217–27.

[44] McNair P, Prapavessis H. Normative data of vertical ground reaction forces during landing from a jump. J Sci Med Sport 1999;2(1):86–8.
[45] Ramey M, Williams K. Ground reaction forces in the triple jump. Int J Sport Biomech 1985;1:233–9.
[46] Avanzini F, Rocchesso D. Modeling collision sounds: non-linear contact force. In: Proceedings of digital audio effects conference; 2001. p. 61–6.
[47] Avanzini F, Rath M, Rocchesso D. Physically-based audio rendering of contact. Proceedings of IEEE international conference on multimedia and expo (ICME'02), vol. 2. IEEE; 2002. p. 445–8.
[48] Adrien J. The missing link: modal synthesis. Cambridge, MA, USA: MIT Press; 1991.
[49] Delle Monache S, Polotti P, Rocchesso D. A toolkit for explorations in sonic interaction design. In: Proceedings of the 5th audio mostly conference. ACM; 2010. p. 1.
[50] Hunt KH, Crossley FRE. Coefficient of restitution interpreted as damping in vibroimpact. ASME J Appl Mech 1975;42(2):440–5.
[51] 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.
[52] Dupont P, Hayward V, Armstrong B, Altpeter F. Single state elastoplastic friction models. IEEE Trans Autom Control 2002;47(5):787–92.
[53] Avanzini F, Rath M, Rocchesso D, Ottaviani L. Low-level models: resonators, interactions, surface textures. In: Rocchesso D, Fontana F, editors. The sounding object, Edizioni mondo estremo; 2003. p. 137–72.
[54] 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.
[55] Moss W, Yeh H, Hong J-M, Lin MC, Manocha D. Sounding liquids: automatic sound synthesis from fluid simulation. ACM Trans Graph 2010;29(3):1–13.
[56] Cirio G, Marchal M, Lécuyer A, Cooperstock J. Vibrotactile rendering of splashing fluids. IEEE Trans Haptics 2013;6(1):117–22.
[57] Drioli C, Rocchesso D. Acoustic rendering of particle-based simulation of liquids in motion. J Multimodal User Interfaces 2012;5(3–4):187–95.
[58] Richardson E. The sounds of impact of a solid on a liquid surface. Proc Phys Soc, Sect B 1955;68(8):541–7.
[59] Leighton T. The acoustic bubble. Academic Press; 1994.
[60] Doel Kvd. Physically based models for liquid sounds. ACM Trans Appl Percept 2005;2(4):534–46.
[61] Longuet-Higgins MS. Nonlinear damping of bubble oscillations by resonant interaction. J Acoust Soc Am 1992;91(3):1414–22.
[62] Minnaert M. On musical air-bubbles and the sounds of running water. Philos Mag 1933;16:235–48.
[63] Freed D. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. J Acoust Soc Am 1990;87(1):311–22.
[64] Giordano B, Petrini K. Hardness recognition in synthetic sounds. In: Proceedings of the Stockholm music acoustics conference, Stockholm, Sweden; 2003.
[65] Rath M, Fontana F. High-level models: bouncing, breaking, rolling, crumpling, pouring. In: Rocchesso D, Fontana F, editors. The sounding object, Edizioni mondo estremo; 2003. p. 111–36.
[66] Alexander R. Principles of animal locomotion. Princeton University Press; 2003.
[67] Nilsson J, Thorstensson A. Adaptability in frequency and amplitude of leg movements during human locomotion at different speeds. Acta Physiol Scand 1987;129(1):107–14.
[68] Young W, Rodger M, Craig C. Perceiving and reenacting spatiotemporal characteristics of walking sounds. J Exp Psychol: Human Percept Perform 2013;39(2):464–76.
[69] Wildes R, Richards W. Recovering material properties from sound. Natural Comput 1988:356–63.
[70] McAdams S, Chaigne A, Roussarie V. The psychomechanics of simulated sound sources: material properties of impacted bars. J Acoust Soc Am 2004;115: 1306–20.
[71] Klatzky R, Pai D, Krotkov E. Perception of material from contact sounds. Presence: Teleoperat Virtual Environ 2000;9(4):399–410.
[72] Avanzini F, Rocchesso D. Controlling material properties in physical models of sounding objects. In: Proceedings of the international computer music conference; 2001. p. 91–4.
[73] Giordano B, McAdams S. Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates. J Acoust Soc Am 2006;119(2):1171–81.
[74] Aramaki M, Gondre C, Kronland-Martinet R, Voinier T, Ystad S. Thinking the sounds: an intuitive control of an impact sound synthesizer. In: Proceedings of the international conference on auditory display; 2009. p. 119–24.
[75] Zanotto D, Turchet L, Boggs E, Agrawal S. Solesound: towards a novel portable system for audio-tactile underfoot feedback. In: Proceedings of the 5th IEEE international conference on biomedical robotics and biomechatronics. IEEE; 2014. p. 193–8.
[76] Harker A, Tremblay P. The hisstools impulse response toolbox: convolution for the masses. In: Proceedings of the international computer music conference. The International Computer Music Association; 2012. p. 148–55.
[77] Visell Y, Law A, Cooperstock JR. Touch is everywhere: floor surfaces as ambient haptic interfaces. IEEE Trans Haptics 2009;2:148–59.
[78] Turchet L. Designing presence for real locomotion in immersive virtual environments: an affordance-based experiential approach. Virtual Real, submitted for publication.
[79] Turchet L, Burelli P, Serafin S. Haptic feedback for enhancing realism of walking simulations. IEEE Trans Haptics 2013;6(1):35–45.
[80] Turchet L, Serafin S. Semantic congruence in audio-haptic simulation of footsteps. Appl Acoust 2014;75(1):59–66.
[81] Leplâtre G, McGregor I. How to tackle auditory interface aesthetics: discussion and case study. In: Proceedings of the international conference on auditory display, vol. 4; 2004. p. 6–9.
[82] Barrass S, Vickers P. Sonification design and aesthetics. In: Hermann T, Hunt A, Neuhoff J, editors. The sonification handbook. Logos Verlag; 2011. p. 145–72.
[83] Nordahl R, Serafin S, Turchet L, Nilsson N. A multimodal architecture for simulating natural interactive walking in virtual environments. PsychNol J 2012;9(3):245–68.
[84] Turchet L, Serafin S, Cesari P. Walking pace affected by interactive sounds simulating stepping on different terrains. ACM Trans Appl Percept 2013;10(4): 23:1–23:14.
[85] Turchet L, Camponogara I, Cesari P. Interactive footstep sounds modulate the perceptual-motor aftereffect of treadmill walking. Exp Brain Res 2015;233: 205–14.
[86] Hermann T, Hunt A, Neuhoff J, editors. The sonification handbook. Logos Publishing House; 2011.
[87] Giordano B, Egermann H, Bresin R. The production and perception of emotionally expressive walking sounds: similarities between musical performance and everyday motor activity. PLOS ONE 2014;9(12):e115,587.
[88] Juslin P, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? Psychol Bull 2003;129(5): 770–814.
[89] Møller H. Fundamentals of binaural technology. Appl Acoust 1992;36(3): 171–218.
[90] Gerzon M. Ambisonics in multichannel broadcasting and video. J Audio Eng Soc 1985;33(11):859–71.
[91] Schacher J, Neukom M. Ambisonics spatialization tools for Max/MSP. In: Proceedings of the international computer music conference; 2006.