What do your footsteps sound like? An investigation on interactive footstep sounds adjustment

Luca Turchet [a,*], David Moffat [b], Ana Tajadura-Jiménez [c], Joshua D. Reiss [b], Tony Stockman [b]

[a] Department of Architecture, Design and Media Technology, Aalborg University Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen, Denmark
[b] Centre for Digital Music, School of Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom
[c] Universidad Loyola Andalucia and UCL Interaction Centre, University College London, 66-72 Gower Street, London WC1E 6EA, United Kingdom

ABSTRACT

This paper presents an experiment where participants were asked to adjust, while walking, the spectral content and the amplitude of synthetic footstep sounds in order to match the sounds of their own footsteps. The sounds were interactively generated by means of a shoe-based system capable of tracking footfalls and delivering real-time auditory feedback via headphones. Results allowed identification of the mean value and the range of variation of spectral centroid and peak level of footstep sounds simulating various combinations of shoe type and ground material. Results showed that the effect of ground material on centroid and peak level depended on the type of shoe. Similarly, the effect of shoe type on the two variables depended on the type of ground material. In particular, participants produced greater amplitudes for hard sole shoes than for soft sole shoes in presence of solid surfaces, while similar amplitudes for both types of shoes were found for aggregate, hybrids, and liquids. No significant correlations were found between each of the two acoustic features and participants' body size. This result might be explained by the fact that while adjusting the sounds participants did not primarily focus on the acoustic rendering of their body. In addition, no significant differences were found between the values of the two acoustic features selected by the experimenters and those adjusted by participants. This result can therefore be considered as a measure of the goodness of the design choices to synthesize the involved footstep sounds for a generic walker. More importantly, this study showed that the relationships between the ground-shoes combinations are not changed when participants are actively walking. This represents the first active listening confirmation of this result, which had previously only been shown in passive listening studies. The results of this research can be used to design ecologically-valid auditory rendering of foot-floor interactions in virtual environments.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Designers of ecologically-valid virtual environments constantly seek to improve their technology in order to create experiences in the virtual world that can be as close as possible to those achievable in a real setting [1]. For a virtual environment to be meaningful in the ecological sense, users must be provided with coherent relations between perception and action. This is especially true for one of the most important tasks in immersive virtual reality, navigation involving real walking [2]. Previous studies demonstrated that real walking is the optimal interaction technique for

navigation of immersive virtual environments since it produces a higher sense of immersion, increases naturalness, and improves task performance compared to other solutions [3–7].

One of the most important acoustic signatures associated to walking in virtual environments is that of footstep sounds. Similar to what happens in real settings, these sounds allow one to perceive not only the surrounding environment but also one's own body. In particular, various authors have highlighted the importance of rendering of the virtual body in a way consistent with the user's body in order to produce a strong feeling of body ownership [8,9,1]. Such a feeling plays a relevant role in the user's experience of the sense of presence, that is, the subjective experience of "being there" inside the virtual world [10]. This is particularly true for locomotion-based applications [1].

To achieve compelling simulations of footstep sounds it is fundamental to consider how those sounds are perceived. This can

inform synthesis techniques based on the analysis-by-synthesis method [11,12]. Such a method consists of designing a model by gathering knowledge about it from data collected via measurements, interviews with experts, or generation of hypotheses, and by successively verifying its validity through synthesis (i.e., by implementing the model in a software tool) and psychophysical tests.

Recently, a novel footstep sound synthesizer based on such a method has been developed [13]. The synthesizer is capable of simulating several types of foot-floor interactions (e.g., different types of steps in walking and running or the sliding of the foot on the floor), different types of shoes and ground materials (solid, aggregate, liquid, and hybrids), as well as some anthropometric features of the walkers (i.e., body size and foot length). The ecological validity of the synthesizer was evaluated by means of listening tests that successfully assessed the effectiveness of the proposed techniques. The synthesizer is based on physical, physically informed, and psychologically informed models, whose control has been designed according to results of various studies about footstep sounds perception in real and virtual settings. The synthesizer can be used in conjunction with several locomotion interfaces (for instance, those developed by Turchet [14]), in order to achieve an interactive sonification of foot-floor interactions [15].

Research about auditory perception of both real and synthesized walked-upon ground materials has shown that humans are capable of classifying with high accuracy the material typology, such as solid (e.g., concrete, wood), aggregate (e.g., gravel, snow), liquid (e.g., water) or hybrid (e.g., mud), but are less precise in discriminating materials belonging to the same typology (e.g., wood can be identified as concrete) [16,17,13]. Along the same lines, it has been proven that the hardness of the sole of the shoes is clearly identifiable both in real footstep sounds produced by walking on solid ground materials [18], and in synthesized footstep sounds simulating walking on materials of various typologies [13]. Various studies have also shown that both real and synthesized footstep sounds are effective in conveying information about gender and anthropometric features of the walker [19,18,20,13]. All those studies consistently revealed that the auditory perception of gender, height, and weight of a walker depends primarily on the footstep sound spectral characteristics. Specifically, footstep sounds having spectra with a predominant high frequency component are associated with females and small body sizes, while maleness and big body sizes are related to spectral dominancy of the low frequencies. Furthermore, research has shown that listeners can identify the emotional state of a walker from the content of both real and synthesized footstep sounds and that such an identification depends on the sound intensity and temporal features (average pace, pace irregularity) [18,21].

Interestingly, Tajadura-Jiménez et al. showed that altering in real-time the spectral content of non-synthesized sounds produced while walking is effective in changing one's own perceived body weight and leads to a related gait pattern [22]. Specifically, their results showed that augmenting the amount of high frequency components of footstep sound not only led to the perception of having a thinner body, but also enhanced the motivation for physical activity, inducing a more dynamic swing and a shorter heel strike. In addition, the increment of high frequency components caused participants to feel more aroused and positive. In a different vein, providing synthetic footstep sounds simulating a surface material having a degree of compliance different from that of the walked-upon one has been proven to affect the locomotion pace of subjects naturally walking at a self-selected speed, as well as to alter the perception of effort and to induce the sensation of sinking into the ground [23].

Following the tenets of the analysis-by-synthesis method, a relevant research question that can inform the synthesis models in

order to achieve better simulations and consistent with users' expectations is how sounds are manipulated when subjects are asked to render their own footstep sounds. This is especially important in interactive scenarios where real walking is involved. While prior studies in virtual settings explored the effects of passive listening to synthesized footstep sounds of an unknown walker, limited research has been conducted so far on the interactive adjustment of the parameters of self-produced synthesized footstep sounds. An exception is the work reported in [24] that faced the issue arising when designing interactive sound rendering for virtual environments, about how to choose the amplitude at which the footstep sounds are delivered. Subjects were asked to adjust the amplitude of interactive footstep sounds on various ground materials, generated by means of a shoe-based system, to the extent they felt appropriate for each of those sounds. Results allowed one to identify for each synthesized material the amplitude mean value and range of variation that could be considered appropriate for different subjects. However, that study did not take into account the spectral properties of the sounds during the adjustment task, nor did the analysis consider eventual correlations between the participants' choices of amplitude and their anthropometric features. From the studies reviewed above it emerges that amplitude and spectral centroid are two parameters of footstep sounds that can be related to the auditory perception of body properties and emotional state, as well as to the rendering of various types of materials and shoes.

In this paper we present an experiment where we asked participants to adjust the spectral content and the amplitude of synthetic footstep sounds in order to match the sounds of their own footsteps. The sounds were provided interactively by means of a shoe-based system capable of tracking footfalls and delivering real-time auditory feedback. Our main goals were: interactively validate the design choices of centroid and peak level for different combinations of ground material and shoe type; assess whether and how these parameters change in relation to actual and perceived body properties; identify the mean values and ranges of variation of the two parameters. This research aimed also at assessing to what extent participants consciously focused on their body properties during their choices of the sounds' parameters. For this purpose we provided participants with rating scales to compare the criteria on which participants relied to perform their choices.

Based on the results about the auditory perception of sole hardness in footstep sounds reported by Giordano et al. [18] and Turchet [13], we hypothesized that participants would have produced greater amplitudes for hard sole shoes than for soft sole shoes in presence of solid surfaces, while similar amplitudes for both types of shoes would have been found for aggregate, hybrids, and liquids. In addition, given that previous works showed how shifting the spectral content of walking sounds influences not only the perceived body size of a heard walker [19,18,20,13], but also the self body size perception [22], we expected a negative correlation between participants' size and spectral centroid. Along the same lines, we hypothesized a positive correlation between participants' size and participants' choices of the sounds' amplitude since the heavier and taller the person, the greater the ground reaction force, i.e., the reaction force produced by the ground at every step [25], and consequently the louder the sound.

With this production-based experimental design we aimed at assessing the goodness of the sound design choices for a generic walker reported in [13]. The footstep sounds involved in the experiment were designed to simulate a genderless walker with a medium body size. The absence of the hypothesized strong correlations between participants' body size and centroid and peak level, as well as the absence of a statistical difference between the participants' and experimenters' choices of the two parameters would

mean that the designed sounds are appropriate for the simulated materials and shoes, and for a generic walker.

We focused on the two acoustic features centroid and peak level because of their importance in footstep sounds perception research, especially for what concerns the auditory perception of the walker's body size. Knowing that subjects tune those parameters to specifically reflect their body size would allow one to inform footstep sounds synthesizers about how to control their parameters to render the walker's virtual body at auditory level [1]. The results of this research are important to inform the design of auditory foot-floor interactions in virtual environments towards more ecologically-valid simulations [26–29].

2. Method

2.1. Participants

Nineteen participants, 12 males and 7 females, aged between 20 and 41 (mean = 28.58, SD = 6.27), took part in the experiment. All participants reported normal hearing and no locomotion impairments. The procedure, approved by the local ethics committee, was in accordance with the ethical standards of the 1964 Declaration of Helsinki. The average duration of the experiment was about one hour. Fatigue was not an issue.

2.2. Apparatus

The apparatus consisted of a pair of sandals (41 size EUR) augmented with pressure sensors [14], a box, attached to a belt, containing an x-OSC wireless micro-controller board, a wired closed headphone set (Sennheiser HD-25-1 II), a MIDI controller equipped with sliders (Korg nanoKONTROL2), and a laptop that ran a Max/MSP[1] application. The latter consisted of a receiver for the MIDI data provided by the MIDI controller, a receiver for the data of the micro-controller board streamed according to the OSC protocol,[2] and the footstep sounds synthesis engine described in [13]. The sandals' shape was adjustable so that it fitted a large range of participants' feet size. A pressure sensor was placed under the sole of each sandal at the level of the heel. The sensors detected feet pressure during contact with the ground; their analog signals were digitized and wirelessly transmitted to the laptop by means of the x-OSC board and used to drive the footstep sound synthesis engine. The synthesized auditory feedback was then conveyed to the user by means of the headphones. Although both the MIDI controller and headphones were connected to the laptop by a wire, while walking participants were barely aware of the presence of the wires since the equipment was light, comfortable and did not constitute any major constraint to their movements. The light box containing the x-OSC board was hung on the back of the user's trousers by means of the belt. The wires coming out from the shoes and directed to the x-OSC board were attached to the user's trousers by means of a tape and secured to the external side of the lower limbs. The USB cable of the MIDI controller was tied together with the wire of the headphones, which was also connected to the laptop. The wires were long enough (5 meters) to allow the participant to move freely. The total latency between the actual footstep fall and the heard synthesized sound was not noticeable: it amounted to about 8 ms (3 ms for the data acquisition and wireless transmission [30], 1 ms for the real-time data analysis, and 4 ms for the auditory feedback synthesis and delivery).

2.3. Stimuli

Stimuli consisted of twelve combinations of ground materials (wood, concrete, gravel, deep snow, mud, and puddle of water) and shoe types (dress shoes and sneakers). Those combinations were chosen because they constituted a comprehensive palette of footstep sounds, involving types of shoes with both hard and soft soles, as well as ground materials belonging to solid, aggregate, liquid, and hybrid typologies. The selection of such stimuli was also inspired by the evaluation of the synthesis engine, which showed that these combinations of simulated ground materials and shoe types were easily recognizable [13]. These material-shoe combinations were also selected because the resulting twelve signals had different features in terms of duration, amplitude, temporal evolution, and spectrum (see Fig. 1). The body size of the walker portrayed by the stimuli was one that could be considered as medium (i.e., not too big nor too small) and of a genderless walker according to the results presented in [13].

The synthesis engine takes into account the fact that in real life, the sound of each step is different from the previous one, as well as the results presented in [31], which demonstrated that the concatenation of the same footstep sound in sequences of walking sounds is perceived as mechanical. Therefore, in order to increase the perceived realism of the interaction and to create sounds more valid from the ecological standpoint [26–29] each sound corresponding to a footfall was simulated in a different way. Specifically, this was achieved by generating for each step a different type of exciter for the synthesis models and a different set of appropriate models' parameters produced with random calculations.

Table 1 shows the features of three footstep sounds corresponding to the twelve stimuli involved in the experiment. The sounds were randomly selected among those generated by the synthesis engine, so can be considered as representative of all the pre-filtered sounds that were actually provided to participants during the experiment. They were used for the analysis of the results (see Section 2.5).

During the experiment participants used five sliders of the MIDI controller to continuously vary five parameters of a two-band parametric equalizer: the global gain, the cut-off frequency and gain of a low-shelf filter, and the cut-off frequency and gain of a high-shelf filter. The range of variation for each of those parameters are reported in Table 2 for each stimulus. Such ranges were chosen in order to generate, for each stimulus, a sound that could be still perceived as representing a simulation of the material-shoe combination. Those ranges were checked in an informal session in which the authors manipulated the five parameters of the experiment stimuli.

2.4. Procedure

The experiment was conducted in an acoustically isolated laboratory. Participants were presented with written instructions. Participants were asked to put on the shoes and headphones described in Section 2.2 and to hold the MIDI controller in one hand while with the other could move the sliders while walking. The task consisted in adjusting the five parameters of the two-band parametric equalizer in order to match the sound that they would actually produce when walking on a specific ground with a specific type of shoe. Participants were informed of the type of material-shoe combination sonically simulated before the beginning of each trial. Notably, we deliberately avoided asking participants to render their own body size because this might have influenced their choices.

They were instructed to walk normally with their natural gait and that the sandals would not track and render the dynamic of the foot-floor interaction. The sandals were not equipped with this

---

[1] https://cycling74.com/.
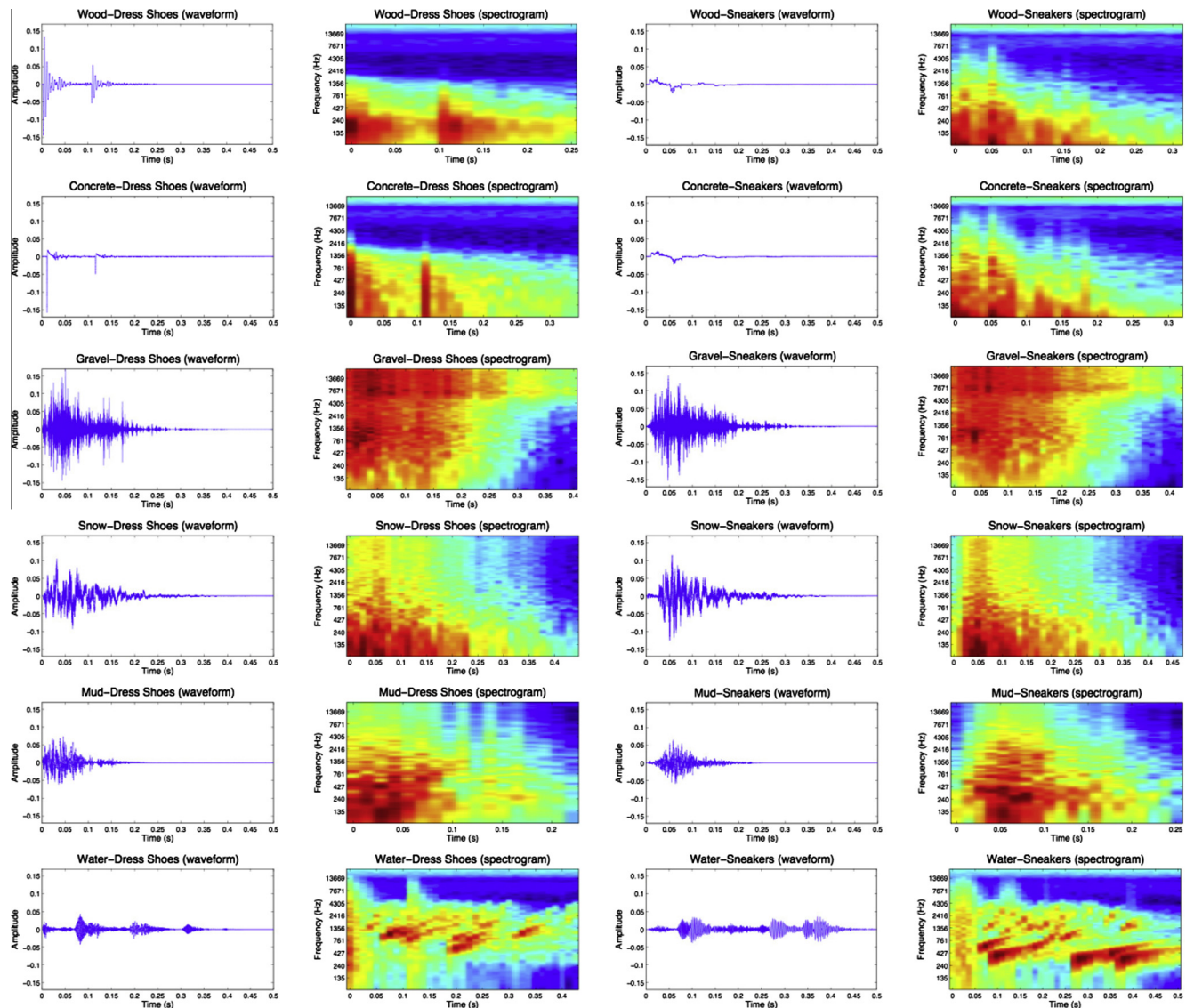[2] http://opensoundcontrol.org/.

Fig. 1. Typical waveforms and spectrogram (with log frequency axis) of the twelve combinations of materials and shoe types used in the experiment.

feature, differently from a system we previously developed [14], since we needed to ensure for each step the same dynamic due to the fact that participants had to adjust the amplitude of the sound. The sounds provided through the headphones completely masked the actual sounds produced by the steps on the laboratory surface. No other instruction was given.

Each participant underwent twelve conditions, each representing a different sound stimulus. Each condition was repeated twice, in a randomized order, for a total of twenty-four trials. In order to avoid biases in participants' evaluations due to the sliders position at the beginning of each trial, we defined two types of adjustment of the filter parameters: the first consisted in mapping the position of the sliders to the ranges indicated in Table 2 from the minimum to the maximum value, the second from the maximum to the minimum value (whether it was the first or the second was counterbalanced across the two repetitions).

Before performing the experiment participants were presented with four practice trials in order to become familiar with the system and with the task. For this purpose, the forest underbrush and marble materials were chosen, each combined with dress shoes and sneakers, as well as with the two different adjustment

types. Those materials were not among those involved in the experiment.

After the walking experiment subjects were asked to evaluate each of the following questions on a visual analog scale (VAS):

Ground. To what extent did you change the parameters of the sound in order to match the specific ground material? [0 = not at all, 10 = very much].

Shoe. To what extent did you change the parameters of the sound in order to match the specific shoe type? [0 = not at all, 10 = very much].

Weight. To what extent did you change the parameters of the sound in order to match your own body weight? [0 = not at all, 10 = very much].

Height. To what extent did you change the parameters of the sound in order to match your own body height? [0 = not at all, 10 = very much].

The order of presentation of the questions was randomized using a 4×4 Latin square. The reason to adopt those questions was to assess whether participants focused on their anthropometric

Table 1
The considered features of three steps randomly chosen for each stimulus.

Stimulus	Step 1		Step 2		Step 3	
	Centroid (Hz)	Peak level (dB)	Centroid (Hz)	Peak level (dB)	Centroid (Hz)	Peak level (dB)
Wood-Dress Shoes	777	−16.75	794	−16.76	825	−16.76
Wood-Sneakers	1878	−32.32	1739	−34.81	1609	−33.13
Concrete-Dress Shoes	1145	−16.06	1121	−16.07	1159	−16.06
Concrete-Sneakers	2093	−33.35	1918	−33.93	1883	−34.76
Snow-Dress Shoes	6085	−19.58	6326	−17.57	6183	−18.97
Snow-Sneakers	6209	−18.1	6317	−19.01	6325	−22.67
Gravel-Dress Shoes	10,351	−15.36	10,509	−14.79	10,408	−13.58
Gravel-Sneakers	10,473	−12.11	10,274	−16.35	10,419	−13.51
Mud-Dress Shoes	4608	−22.62	4729	−20.79	5031	−26.69
Mud-Sneakers	4521	−10.79	4670	−23.76	4653	−23.81
Water-Dress Shoes	1905	−27.33	1040	−20.84	1406	−26.59
Water-Sneakers	1462	−28.3	935	−19.08	1378	−27.78

Table 2
Range of variation of the five parameter of the two-band parametric equalizer for each stimulus.

Stimulus	Global gain (dB)	Low-shelf cut-off (Hz)	Low-shelf cut-off gain (dB)	High-shelf cut-off (Hz)	High-shelf cut-off gain (dB)
Wood-Dress Shoes	[−12, 12]	[150, 500]	[−24, 24]	[200, 700]	[−24, 24]
Wood-Sneakers	[−12, 12]	[150, 500]	[−24, 24]	[200, 700]	[−24, 24]
Concrete-Dress Shoes	[−12, 12]	[200, 600]	[−24, 24]	[300, 1000]	[−24, 24]
Concrete-Sneakers	[−12, 12]	[200, 600]	[−24, 24]	[300, 1000]	[−24, 24]
Gravel-Dress Shoes	[−12, 12]	[200, 600]	[−24, 24]	[1200, 3500]	[−24, 24]
Gravel-Sneakers	[−12, 12]	[200, 600]	[−24, 24]	[1200, 3500]	[−24, 24]
Snow-Dress Shoes	[−12, 12]	[200, 600]	[−24, 24]	[350, 800]	[−24, 24]
Snow-Sneakers	[−12, 12]	[200, 600]	[−24, 24]	[350, 800]	[−24, 24]
Mud-Dress Shoes	[−12, 12]	[100, 600]	[−24, 24]	[200, 800]	[−24, 24]
Mud-Sneakers	[−12, 12]	[100, 600]	[−24, 24]	[200, 800]	[−24, 24]
Water-Dress Shoes	[−12, 12]	[450, 800]	[−24, 24]	[500, 3000]	[−24, 24]
Water-Sneakers	[−12, 12]	[450, 800]	[−24, 24]	[500, 3000]	[−24, 24]

features rather than other aspects while rendering their own footstep sounds.

Subsequently, participants' age, weight, height, and shoe size were collected. Weight and height were measured by means of a scale and a meter respectively, but participants were not informed about the actual measurements. This was in order to avoid possible biases in the immediately subsequent task that assessed their perceived body weight and height. For this purpose we used a body visualization tool [3] (see Fig. 2) adopted by other studies for the same purpose [32,22]. In two trials, participants adjusted the weight and height related dimension of the body of a 3D avatar displayed on the screen to correspond to their own perceived body dimensions [33,34,32,22]. The values of the two parameters were hidden by means of a mask placed on the computer screen, so that only the corresponding sliders could be seen. The mask also covered all the other parameters of the tool (those indicated in blue in Fig. 2). While the avatar's parameters were set to match gender of each participant, the initial avatar's weight and height varied across trials to avoid anchor effects of the initial value [33,34,32,22]. This was set to match the participant's weight ±25% and height ±15% (whether it was + or − was counterbalanced across the two trials).

2.5. Data handling

For each participant and for each trial, the values of the five parameters of the two-band parametric equalizer produced by the selection of the five sliders positions were recorded. Three audio files were created, for each trial and for each participant, as a result of the filtering the three footstep sounds presented in Table 1. These files were generated offline by the synthesis engine set to simulate the twelve stimuli. A total of 1368 files (mono, sample rate = 44,100 Hz, bit depth = 16) was created and analyzed.

Two features were extracted from each of the generated files containing the filtered footstep sounds: centroid and peak level. Those features were selected as they are representative of the spectral and temporal content of the stimuli respectively, and were found to be related to perception of body size in previous research on synthesized footstep sounds [13]. The spectral centroid was calculated using the MIRToolbox 1.5 [35]. The frame length to calculate the two features was identical for each file resulting from filtering the files belonging to a given stimulus. It consisted by considering as beginning and end of a footstep the nearest zero crossings above a given threshold.

3. Results

Table 3 reports the means, standard deviation, and confidence intervals for centroid and peak level, averaged for all participants. A Pearson correlation coefficient was computed to assess the relationship between the two parameters. Results showed a weak correlation between them ($r(1366) = 0.13$, $p < .001$).

Statistical analysis was performed by means of two-way repeated measures ANOVAs, by considering the six ground materials and the two types of shoe for each of the two dependent variables (centroid, peak level, and loudness).[4] For each of the ANOVAs, the results of the Mauchly's tests revealed that the assumption of sphericity was met. All post hoc analyses were performed by using Tukey's procedure. In presence of significant interaction effects, we present the results of separate one-way ANOVAs (corrected with Bonferroni's correction) for both comparisons between the levels of shoe type within each level of ground, and between the levels of ground within each level of shoe type [36].

---

[3] http://www.bodyvisualizer.com.

[4] A two-way repeated measures MANOVA on the two selected variables could not be performed due to the small sample size involved in this study.
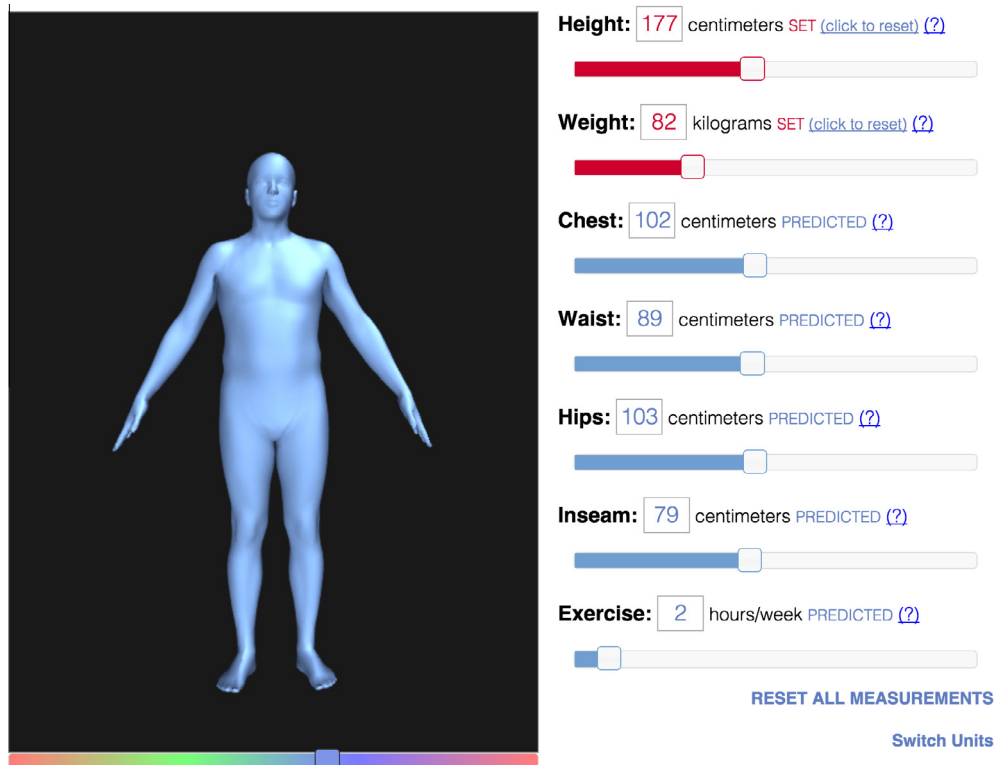
Fig. 2. A screenshot of the body visualization tool involved in the experiment (taken from http://www.bodyvisualizer.com).

Table 3
Means, standard deviations and confidence intervals for the two parameters centroid and peak level as rendered by participants for the twelve combinations of ground material and shoe type.

Stimulus	Centroid (Hz)			Peak level (dB)		
	Mean	Std. Dev.	95% CI	Mean	Std. Dev.	95% CI
Wood-Dress Shoes	2058	1324	[1812, 2303]	−20.09	8.47	[−21.67, −18.52]
Wood-Sneakers	2294	1875	[1946, 2642]	−27.70	10.57	[−29.66, −25.74]
Concrete-Dress Shoes	2057	1655	[1750, 2364]	−20.99	10.11	[−22.86, −19.11]
Concrete-Sneakers	2876	1493	[2599, 3153]	−34.87	11.35	[−36.98, −32.77]
Gravel-Dress Shoes	9554	1617	[9254, 9854]	−17.14	7.54	[−18.54, −15.74]
Gravel-Sneakers	8769	1688	[8456, 9082]	−20.33	7.58	[−21.74, −18.92]
Snow-Dress Shoes	6373	1576	[6081, 6666]	−25.22	10.28	[−27.13, −23.31]
Snow-Sneakers	5776	1715	[5458, 6094]	−26.61	9.83	[−28.44, −24.79]
Mud-Dress Shoes	4208	1600	[3911, 4505]	−26.61	9.28	[−28.33, −24.89]
Mud-Sneakers	3765	1180	[3546, 3984]	−27.05	12.56	[−29.38, −24.72]
Water-Dress Shoes	1507	1111	[1301, 1713]	−28.10	10.78	[−30.10, −26.09]
Water-Sneakers	1139	737	[1002, 1276]	−28.32	10.84	[−30.33, −26.31]

With regards to the centroid, the ANOVA yielded a significant main effect for material, $F(5,90) = 191.7$, $p < .001$, as well as for the interaction effect between material and shoe type, $F(5,90) = 4.835$, $p < .001$. No significant main effect of shoe type was found. Pairwise comparisons for material are shown in Table 4. The effect of ground material on centroid depended on the type of shoes, as evidenced by the significant interaction between shoe type and ground material (see Fig. 3). Separate one-way ANOVAs showed that the effect of ground material was significant for both dress shoes and sneakers ($F(5,90) = 156$, $p < .001$ and $F(5,90) = 118.4$, $p < .001$ respectively). Comparisons for each pair of material and for both shoes are shown in Table 5. Symmetrically, the effect of shoe type on centroid depended on the type of ground material. Separate one-way ANOVAs showed that the centroid was greater for sneakers compared to dress shoes for the concrete material ($F(1,18) = 9.623$, $p < .01$),

As far as the peak level is concerned, the ANOVA yielded a significant main effect for shoe, $F(1,18) = 58.12$, $p < .001$, for material, $F(5,90) = 7.76$, $p < .001$, as well as for their interaction effect $F(5,90) = 12.21$, $p < .001$. Pairwise comparisons for material are shown in Table 4. The effect of ground material on peak level depended on the type of shoes, as evidenced by the significant interaction between shoe type and ground material (see Fig. 3). Separate one-way ANOVAs showed that the effect of ground material was significant for both dress shoes and sneakers ($F(5,90) = 9.831$, $p < .001$ and $F(5,90) = 8.309$, $p < .001$ respectively). Comparisons for each pair of material and for both shoes are shown in Table 5. Symmetrically, the effect of shoe type on peak level depended on the type of ground material. Separate one-way ANOVAs showed that peak level of wood and concrete was significantly greater for dress shoes compared to sneakers (respectively $F(1,18) = 19.87$, $p < .01$ and $F(1,18) = 82.09$, $p < .001$).

Table 4
Results of the pairwise comparisons for ground material on the two investigated features.

Pair	Centroid	Peak level
Gravel-Concrete	***	***
Mud-Concrete	***	–
Snow-Concrete	***	–
Water-Concrete	***	–
Wood-Concrete	–	***
Mud-Gravel	***	***
Snow-Gravel	***	***
Water-Gravel	***	***
Wood-Gravel	***	***
Snow-Mud	***	–
Water-Mud	***	–
Wood-Mud	***	***
Water-Snow	***	*
Wood-Snow	***	–
Wood-Water	***	***

Legend: * represents $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and $-p \geqslant 0.05$.

To search for correlations between each of the two investigated parameters and participants' body size as a function of the combination of ground material and shoe type, we performed linear mixed-effects models analyses. Since we found a high correlation between actual height and weight ($r(1366) = 0.82$, $p < .001$), we did not include both of them in order to not violate the assumption of absence of collinearity required by the analysis. Similarly, a high correlation were found between perceived height and weight ($r(1366) = 0.84, p < .001$). Therefore, we selected actual weight and perceived weight, and we discarded actual height and perceived height. We used two models for each of the two dependent variables centroid and the peak level. In a first model, as random effects, we had intercepts for subjects, as fixed effects we had actual weight, ground material, shoe type, and their interaction. In a second model, as random effects, we had intercepts for subjects, as fixed effects we had perceived weight, ground material, shoe type, and their interaction. Results showed that in none of the performed analyses the correlations between either actual or perceived weight and each of the two investigated parameters were significant.

In order to assess the degree of similarity between experimenters' choices (reported in Table 1) and participants' adjustments (reported in Table 3) we performed two-way repeated measures ANOVAs for each of the dependent variables centroid and peak level, having as factors the type of stimulus (twelve levels, i.e., the twelve ground-shoe combinations) and the type of parameter values (two levels, i.e., designed by the experimenter or adjusted by participants). In both analyses, results showed a significant main effect for type of stimulus ($F(11,198) = 131.307$, $p < .001$ for centroid, $F(11,198) = 11.528$, $p < .001$ and for peak level), while the effect of type of parameter values and the interaction effect were not significant.

Finally, we analyzed the participants' evaluations of the rating scales described in Section 2.4 by means of several paired t-tests with Bonferroni's correction. Fig. 4 shows the evaluations expressed as VAS score for each of the four questions. Results revealed that participants focused on the rendering of the ground material to a significantly greater extent compared to weight ($p < 0.01$) and height ($p < 0.001$); they focused on the rendering of the shoe type to a significantly greater extent compared to height ($p < 0.001$); and they focused on the rendering of the weight to a significantly greater extent compared to height ($p < 0.05$).

4. Discussion

Results, illustrated in Fig. 3 and Table 5, revealed that the effect of ground material on centroid and peak level depended on the type of shoes. Similarly, the effect of shoe type on the two variables was dependent on the type of ground material. From those results it can also be seen that the centroid of wood was very similar to that of concrete in the presence of the dress shoes condition, while, consistently with the design for those sounds reported in [13], it was significantly lower for wood compared to concrete in the presence of the sneakers condition. The absence of a significant difference between those two ground materials in the presence of the dress shoes parallels the findings on the confusion in the identification performances of both real and synthesized footstep sounds reported for materials belonging to the solid typology impacted by a hard sole shoe type [16,17,13]. Consistently with what is suggested by the common everyday experience and in agreement with our hypothesis, participants set the amplitude of the solid materials impacted by hard sole shoes to values higher than those of the same materials impacted by soft sole shoes. As expected, for all the other aggregate, liquid and hybrid materials involved the chosen amplitudes did not depend from the type of shoe.

From Table 3, it can be seen that there was a high variability in the participants' choices of the two variables. On the one hand, this can be explained in part by the fact that in real life each of the ground materials simulated in our study comes with various types (e.g., thicker or thinner grain sizes in gravel, different types of stiffness of wood, etc.), as well as the soles of the shoes. Those differences in the properties of the impacting and impacted materials can result in rather different footstep sounds for a given combination of shoe and ground. On the other hand, an explanation can be attributed to the different degrees of familiarity that each participant had with the sounds and to the fact that each participant had his/her own mental model of each combination of shoe and ground material, which is related to ecological frequency as demonstrated by Ballas [37].

Contrarily to our hypotheses, participants did not adjust either centroid or peak level according to their own body size. This result is not in agreement with previous findings on footstep sounds research in both real and virtual settings [19,18,22,20,13]. As a consequence, our result did not allow to identify the optimal amplitude and centroid for rendering a given body size. However, it would be not appropriate to conclude, based on these results only, that the rendering of the walkers' body size by means of footstep sounds is not a relevant aspect. Probably the fact that the correlations between each of the two parameters and body dimensions was not found can be explained by the results of the evaluations on the four rating scales (see Fig. 4). From such results it clearly emerged that participants did not consciously adjusted the sound amplitude and spectral content to specifically match their own body size, but they focused more on the rendering of the surface material and of the shoe type. Therefore, more research is needed to allow us to identify the best simulations of each shoe-ground combination for the rendering of the virtual body. As suggested by Turchet [1], when the footstep sounds provided interactively are consistent with the user's actual body size, they could lead to stronger feelings of body ownership and as a consequence to a higher sense of presence. Conversely, in the absence of such a consistence, it might be possible to alter the user's body size perception, replicating the results reported by Tajadura-Jiménez et al. [22] for the case of non-synthesized footstep sounds.

Taken together, these results, interactively produced, can be considered as a proof of the goodness of the sound design choices related to the tuning of the footstep sound synthesizer's parameters, which previously were only passively assessed [13]. No significant differences were found between the patterns of centroid and peak level selected by experimenters' and participants' for each combination of shoe type and ground materials. Therefore, the reported results can be applied to tune a footstep sounds synthesizer to render steps of a generic walker for the investigated shoe types and ground materials.
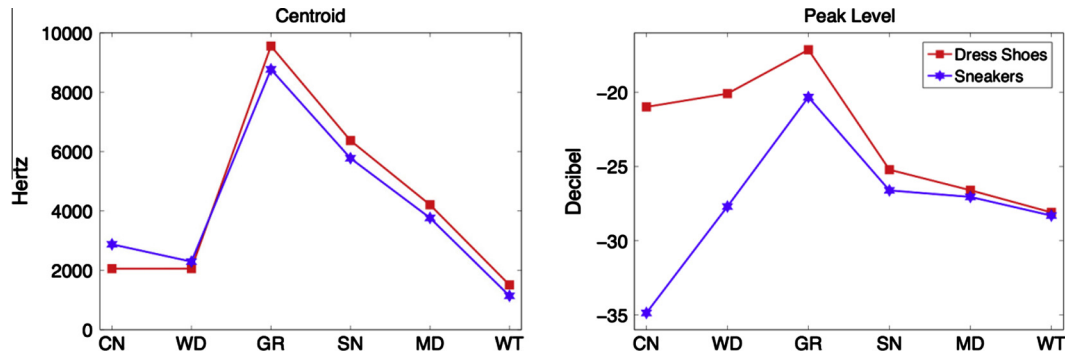
Fig. 3. Graphical representation of the interaction effects between shoe type and material for centroid (left) and peak level (right). Legend: CN = concrete, WD = wood, GR = gravel, SN = snow, MD = mud, WT = water.

Table 5
Results of the pairwise comparisons for ground material on the two investigated features, separately for the two types of shoes.

Pair	Centroid		Peak level	
	Ds	Sn	Ds	Sn
Gravel-Concrete	***	***	***	***
Mud-Concrete	***	***	***	***
Snow-Concrete	***	***	***	***
Water-Concrete	–	***	–	***
Wood-Concrete	–	*	***	***
Mud-Gravel	***	***	***	***
Snow-Gravel	***	***	***	***
Water-Gravel	***	***	***	***
Wood-Gravel	***	***	–	***
Snow-Mud	***	***	–	–
Water-Mud	***	***	–	–
Wood-Mud	***	***	***	–
Water-Snow	***	***	–	–
Wood-Snow	***	***	***	–
Wood-Water	–	***	***	–

Legend: * represents $p < 0.025$ (Bonferroni's corrected alpha level), *** $p < 0.001$, and $-p \geqslant 0.025$. Ds = dress shoes, Sn = sneakers.
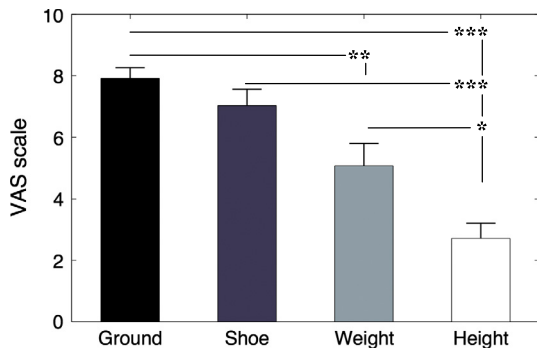

Fig. 4. Graphical representation of the mean and the standard error for participants' evaluations on the investigated rating scales. Legend: * represents $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Interestingly, this result about the similarities between the passive assessment of the footstep sounds and their interactive production parallels the absence of differences between the interactive and passive production of emotional gait styles reported in [21,38] respectively. Those studies involved the same footstep sound synthesizer used here. Together with the present study, they suggest that strong similarities exist between imaginary and real motor activity, as evidenced by various neuroscientific and perceptual studies [39–42].

On a separate note, it is possible that the results of our study could have been affected by the sensory mismatch of wearing

sandals and walking on a solid surface, while hearing simulations of sneakers and dress shoes on solid, hybrid and liquid ground materials. Previous results on perception of synthesized footstep sounds involving the same interactive system revealed the presence of pseudo-haptic illusions [23] and footsteps localization biases [43] that are hypothesized to be due to conflicts between the auditory and the tactile perceptual systems.

Finally, one might consider that to generalize the results of the present study and of the study reported in [13] a comparison with real footstep sounds would be necessary. Although a comprehensive analysis of real footstep sounds is not available (for reasons discussed in Section 2.1 of [13]), some comparisons can be found in manifold studies available in the literature. The study presented in [16] reports an experiment on perception of surface materials in real and synthesized footstep sounds. Results showed that the identification performances in the two conditions were similar. Such results, together with those presented in [13], were also in agreement with those of the study involving surface material perception in real footstep sounds reported in [17]. Furthermore, the study reported in [38] presented a production and identification experiment on emotionally expressive walking styles simulated using the same synthesizer involved in the present study. Results were in agreement with those of a similar experiment involving real footstep sounds, which is reported in [18].Consistently, all such studies reveal that the involved synthesized sounds are identified at a performance level that is comparable with that achievable with real footstep sounds. This is an indication of the ecological validity and success of our simulation algorithms and of their control.

5. Conclusion

In this work we presented an experiment in which participants used a gesture controller for the real-time manipulation of the amplitude and spectral content of synthesized footstep sounds interactively generated by means of sensorized shoes. The manipulations were achieved through five parameters of a two-band equalizer that allowed to vary two acoustical variables (spectral centroid and peak level), which are related to the perception of ground material, shoe, and body size in footstep sounds. This approach, based on the analysis-by-synthesis method [11,12], resulted in the identification of mean values and range of variation of those variables.

The main result of this study is that the production experiment confirmed the results from the passive listening experiment reported in [13]. No significant differences were found between the values of the two acoustic features selected by the experimenters and those adjusted by participants. This result can therefore be considered as a measure of the goodness of the design

choices to synthesize the involved footstep sounds for a generic walker.

Results showed that the effect of ground material on centroid and peak level depended on the type of shoe. Similarly, the effect of shoe type on the two variables depended on the type of ground material. In particular, participants produced greater amplitudes for hard sole shoes than for soft sole shoes in presence of solid surfaces, while similar amplitudes for both types of shoes were found for aggregate, hybrids, and liquids.

Contrarily to our hypothesis, we did not find significant correlations between each of the two acoustic features and participants' body size. However, this result might be explained by the fact that while adjusting the sounds participants did not primarily focus on the acoustic rendering of their body.

From the applicative standpoint, these results can be used for both the design and control of synthesized footstep sounds that are more valid from the ecological standpoint [26–29]. Indeed, the mean values and confidence intervals reported in Table 3 can be practically utilized by sound designers to create footstep sounds for each of the reported ground material-shoe type combinations and that fit a generic body size.

In future work we plan to investigate further the relationship between walkers' body size and the two investigated acoustic features in order to assess whether optimal values can be found for a given body size for various combinations of shoe type and ground material. Along the same line, we plan to conduct an experiment similar to the one here presented, where participants can control other aspects of the sound, (e.g., attack type, duration). Such research could contribute to the design of better interfaces for the interactive sonification of foot-floor interactions. Moreover, we plan to assess the effects of the resulting simulations on sense of agency, body size perception and gait modulation.

Acknowledgments

References

[1] Turchet L. Designing presence for real locomotion in immersive virtual environments: an affordance-based experiential approach. Virt Real 2015;19(3):277–90.

[2] Steinicke F, Visell Y, Campos J, Lécuyer A. Human walking in virtual environments: perception, technology, and applications. Springer Verlag; 2013.

[3] Slater M, Usoh M, Steed A. Taking steps: the influence of a walking technique on presence in virtual reality. ACM Trans Comput-Hum Interact 1995;2(3):201–19.

[4] Usoh M, Arthur K, Whitton M, Bastos R, Steed A, Slater M, et al. Walking > walking-in-place > flying, in virtual environments. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques. p. 359–64.

[5] Zanbaka C, Lok B, Babu S, Ulinski A, Hodges L. Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. IEEE Trans Vis Comput Graph 2005;11(6):694–705.

[6] Ruddle R, Lessels S. The benefits of using a walking interface to navigate virtual environments. ACM Trans Comput-Hum Interact 2009;16(1):5.

[7] Peck T, Fuchs H, Whitton M. The design and evaluation of a large-scale real-walking locomotion interface. IEEE Trans Vis Comput Graph 2012;18(7):1053–67.

[8] Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans Roy Soc B: Biol Sci 2009;364(1535):3549–57.

[9] Spanlang B, Normand J, Borland D, Kilteni K, Giannopoulos E, Pomes A, et al. How to build an embodiment lab: achieving body representation illusions in virtual reality. Front Robot AI 2014;1:9.

[10] Heeter C. Being there: the subjective experience of presence. Presence: Teleop Virt Environ 1992;1(2):262–71.

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

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

[13] Turchet L. Footstep sounds synthesis: design, implementation, and evaluation of foot-floor interactions, surface materials, shoe types, and walkers' features. Appl Acoust 2016;107:46–68.

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

[15] Hermann T, Hunt A, Neuhoff J, editors. The sonification handbook. Logos Publishing House; 2011.

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

[17] 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.

[18] 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):e115587.

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

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

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

[22] Tajadura-Jiménez A, Basia M, Deroy O, Fairhurst M, Marquardt N, Bianchi-Berthouze N. As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. ACM; 2015. p. 2943–52.

[23] 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.

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

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

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

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

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

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

[30] Madgwick S, Mitchell T. x-osc: A versatile wireless i/o device for creative/music applications. In: Proceedings of sound and music computing conference.

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

[32] Piryankova I, Stefanucci J, Romero J, Rosa S, Black M, Mohler B. Can i recognize my body's weight? the influence of shape and texture on the perception of self. ACM Trans Appl Percept 2014;11(3):13.

[33] Cazzato V, Mian E, Serino A, Mele S, Urgesi C. Distinct contributions of extrastriate body area and temporoparietal junction in perceiving ones own and others body. Cogn Affect Behav Neurosci 2014;15(1):211–28.

[34] Legenbauer T, Vocks S, Betz S, Báguena Puigcerver A, Benecke MJ, Troje N, et al. Differences in the nature of body image disturbances between female obese individuals with versus without a comorbid binge eating disorder: an exploratory study including static and dynamic aspects of body image. Behav Modif 2011;35(2):162–86.

[35] Lartillot O, Toiviainen P. MIR in matlab (ii): a toolbox for musical feature extraction from audio. In: Dixon S, Bainbridge D, Typke sR, editors. Proceedings of the 8th international conference on music information retrieval. p. 237–44.

[36] Meyers L, Gamst G, Guarino A. Applied multivariate research: design and interpretation. Sage; 2006.

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

[38] Turchet L, Rodà A. Emotion rendering in auditory simulations of imagined walking styles. IEEE Trans Affect Comput 2016. http://dx.doi.org/10.1109/TAFFC.2016.2520924. in press.

[39] Porro C, Francescato M, Cettolo V, Diamond ME, Baraldi P, Zuiani C, et al. Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. J Neurosci 1996;16(23):7688–98.

[40] Decety J. The neurophysiological basis of motor imagery. Behav Brain Res 1996;77(1):45–52.

[41] Decety J, Jeannerod M, Prablanc C. The timing of mentally represented actions. Behav Brain Res 1989;34(1):35–42.

[42] Young W, Rodger M, Craig C. Perceiving and reenacting spatiotemporal characteristics of walking sounds. J Exp Psychol Hum Percept Perform 2013;39(2):464–76.

[43] Turchet L, Spagnol S, Geronazzo M, Avanzini F. Localization of self generated synthetic footstep sounds on different walked-upon materials through headphones. Virt Real 2016;20(1):1–16.