Virtual Reality (2016) 20:1–16
DOI 10.1007/s10055-015-0272-6

ORIGINAL ARTICLE

Localization of self-generated synthetic footstep sounds on different walked-upon materials through headphones

Luca Turchet[1] · Simone Spagnol[2] · Michele Geronazzo[3] · Federico Avanzini[3]

Received: 10 September 2014 / Accepted: 5 August 2015 / Published online: 21 August 2015

Abstract This paper focuses on the localization of footstep sounds interactively generated during walking and provided through headphones. Three distinct experiments were conducted in a laboratory involving a pair of sandals enhanced with pressure sensors and a footstep synthesizer capable of simulating two typologies of surface materials: solid (e.g., wood) and aggregate (e.g., gravel). Different sound delivery methods (mono, stereo, binaural) as well as several surface materials, in the presence or absence of concurrent contextual auditory information provided as soundscapes, were evaluated in a vertical localization task. Results showed that solid surfaces were localized significantly farther from the walker's feet than the aggregate ones. This effect was independent of the used rendering technique, of the presence of soundscapes, and of merely temporal or spectral attributes of sound. The effect is hypothesized to be due to a semantic conflict between auditory and haptic information such that the higher the semantic incongruence the greater the distance of the perceived sound source from the feet. The presented results contribute to the development of further knowledge toward a basis for the design of continuous multimodal feedback in virtual reality applications.

Keywords Walking · Interactive auditory feedback · Localization

✉ Luca Turchet
tur@create.aau.dk

Simone Spagnol
spagnols@hi.is

Michele Geronazzo
geronazzo@dei.unipd.it

Federico Avanzini
avanzini@dei.unipd.it

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

[2] Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, School of Engineering and Natural Sciences, University of Iceland, Tæknigarður Dunhagi 5, 107 Reykjavík, Iceland

[3] Department of Information Engineering, University of Padova, Via Gradenigo 6/A, 35131 Padua, Italy

1 Introduction

Recent research in the field of multimodal virtual environments has focused on the simulation of foot–floor interactions (Steinicke et al. 2013; Visell et al. 2009) by addressing the problem of enhancing their realism at auditory and haptic levels in order to achieve higher level of presence (Slater et al. 2009; Turchet 2015). As a matter of fact, the human brain relies on inputs from different senses to form a coherent percept of the environment. These pieces of information usually complement and confirm each other, thereby enhancing reliability of percepts (Stein and Meredith 1993).

In particular, several results have indicated that the typology of the surface onto which we walk is processed very consistently in both the auditory and haptic modalities. The excellent somatosensory capacities of the human feet have been demonstrated to be capable of discriminating with high accuracy different types of surfaces (Kobayashi et al. 2008; Giordano et al. 2012). Similarly, studies on the ability to identify ground materials simulated either with auditory or with haptic information (Serafin et al. 2010; Nordahl et al. 2010) revealed that material typology is consistently recognized by using both modalities.

Turchet proposed a footstep sound synthesis engine, based on physical models, which allows the simulation of two typologies of ground materials: solid (i.e., homogeneous floors like wood or metal) and aggregate (i.e., grounds possessing a granular structure like gravel or snow) (Turchet 2015). The ecological validity of such simulations was assessed with experiments in which subjects were asked to recognize the synthesized materials (Nordahl et al. 2010). Results showed that subjects were able to recognize the synthesized surfaces with an accuracy comparable to that of real recorded footstep sounds, which was an indication of the success of the proposed algorithms and their control.

A complicating factor is that various sound reproduction methods can be used to deliver the synthesized sounds to the walker: loudspeakers directly placed on top of the shoes (Papetti et al. 2010; Zanotto et al. 2014), on their soles (Papetti et al. 2011), or embedded in the walking surface (Visell et al. 2008). Also, the interactive delivery of footstep sounds can be achieved by means of a surround sound systems composed of loudspeakers (Turchet and Serafin 2011), while no extensive research has been conducted into headphone-based reproduction of interactive locomotion sounds.

Even more importantly, to our knowledge no previous research has systematically addressed the issue of footstep sound localization in VR contexts. The main goal of this work is thus to investigate the role of auditory information in modulating the localization of self-generated footstep sounds and to test whether differences in perceived localization of footstep sounds affect the realism and naturalness of the walking experience as well as the sense of disorientation associated with different layers of auditory information. To this end, we consider different techniques for footstep sounds rendering by means of headphones, which despite presenting possible disadvantages (e.g., invasiveness), possess a number of desirable features. In particular they eliminate reverberation and other acoustic effects of the real listening space, reduce background noise, and provide adaptable audio displays. More importantly, they allow the delivery of stimuli with different degrees of spatiality, e.g., mono (=0 dimensions), stereo (=1 dimension), and binaural (=2/3 dimensions) reproduction by means of head-related transfer functions (HRTFs) (Cheng and Wakefield 2001). Furthermore, we assess the relative importance of auditory spatial cues with respect to semantic information such as walking surface and context as well as to signal-level features.

The remainder of the paper is organized as follows. Section 2 reports the design and results of experiment 1, whose main goal is to investigate whether different sound rendering techniques have an influence on the localization of solid and aggregate footstep sounds. The role of contextual information (soundscapes) is instead explored in experiment 2, described in Sect. 3. In the final experiment, reported in Sect. 4, we consider a larger sonic palette to test whether signal-level features affect the results found in the previous two experiments. Sections 5 and 6 conclude the paper with a general discussion on the global results of the three experiments and the implications they provide to the design of walking VR experiences.

2 Experiment 1

This first experiment was designed so as to explore whether different audio-rendering techniques over headphones (mono, stereo, binaural) affect localization judgments of synthetic self-generated footstep sounds on four different surface materials simulating two different surface typologies, i.e., aggregate and solid. Such a distinction is motivated by a previous preliminary study (Turchet and Serafin 2011) that highlighted significant differences (in terms of localization, realism, naturalness of the interaction, and sense of disorientation) between the perception of dynamically generated footstep sounds on aggregate and solid surfaces provided via loudspeakers.

The basic idea of the binaural technique is that by recording real-life sounds inside a person's ears, the appropriately post-processed sound file played back through headphones will be perceived by that person almost as realistic as the original one. In order to find the correct sound pressure that an arbitrary source produces at the eardrum, we need the impulse response from the source to the eardrum, called head-related impulse response (HRIR), whose Fourier transform is known as head-related transfer function (HRTF) (Cheng and Wakefield 2001). The HRTF captures all the acoustic cues used for source localization; once the HRTFs for the left and the right ear are known, accurate binaural signals can be generated starting from a monaural sound source.

Our starting hypothesis is that if the footstep sound has sufficient duration and high-frequency content (Vliegen and Opstal 2004; Hebrank and Wright 1974) in order to enable vertical localization mechanisms, which is the case for aggregate surface sounds as opposed to solid surface sounds, then different rendering techniques should result in different localization ratings. In particular, binaural techniques should allow the walker to perceive synthesized aggregate footstep sounds as coming from below, despite the known difficulty in localizing virtual sources near the median plane, with an accuracy that shall depend on the degree of customization of the used HRTFs (Møller et al. 1996; Wenzel et al. 1993). Different localization ratings should in turn modulate the perception of the realism, naturalness, and sense of disorientation of the walking experience.

2.1 Participants

Twelve participants, seven males and five females, aged between 19 and 31 ($M = 22.41, \mathrm{SD} = 4.23$), took part in the experiment. All participants reported normal hearing and no impairment in locomotion.

2.2 Apparatus

The experiment was carried out in a quiet room where the setup was installed, and the walking area was 3.2×2.9 m wide (see Figs. 1, 2). It consisted of a MacBook Pro laptop, running a sound synthesis engine (Turchet 2015); a pair of soft sole sandals enhanced with pressure sensors (placed in correspondence with the heel); an Arduino UNO board, managing the sensors' data acquisition; a Fireface 800 soundcard; a pair of Sennheiser HDA 200 headphones. These headphones were mainly chosen because of their closed form facilitating isolation from external noise and the flatness of their frequency response (Boren et al. 2014).

Footstep sound synthesis was interactively driven during locomotion of the subject wearing the shoes. The description of the control algorithms based on the analysis of the values of the pressure sensors, implemented in Max/MSP, can be found in (Turchet et al. 2010). The generated audio stream was then sent in real time to a Pure Data patch responsible for the different audio-rendering techniques.

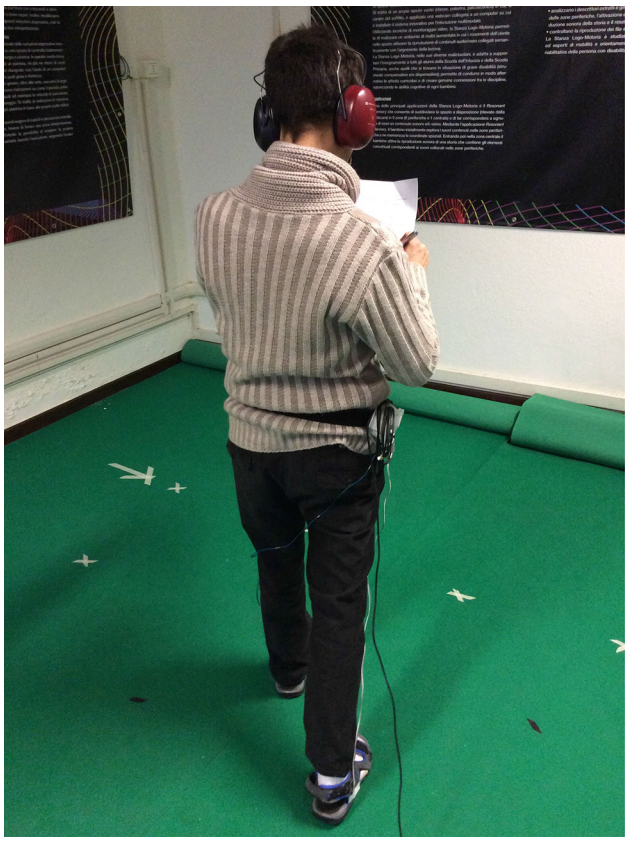

Fig. 1 A subject performing the experiment

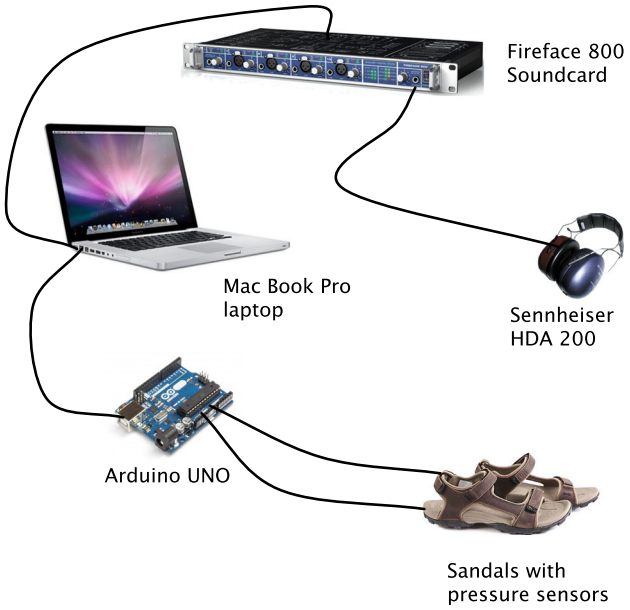

Fig. 2 Block diagram of the interactive system

2.3 Stimuli

The used hardware allowed real-time control of the sound synthesis engine, which was set so as to synthesize footstep sounds on four surface materials: two solid (wood and metal) and two aggregate (snow and gravel).[1]

Solid materials were simulated using an impact model (Avanzini and Rocchesso 2001). In the simulation of impact with solids, the contact was modeled by a Hunt–Crossley-type interaction where the force f between two bodies combines hardening elasticity and a dissipation term (Hunt and Crossley 1975):

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

where x represents contact interpenetration (when $x > 0$ the two objects are in contact), $\dot{x}$ is compression velocity, k accounts for 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. The described model was discretized as proposed by Avanzini and Rocchesso (2001).

To simulate aggregate surfaces, the physically informed sonic models algorithm was adopted (Cook 1997). This

[1] Audio examples of the involved stimuli can be found at http://www.ahws-project.net/audio.html. A video of an apparatus similar to that involved in the experiment can be found at http://www.youtube.com/watch?v=kRKcKgYCPCY.

algorithm simulates particle interactions by using a stochastic parameterization, thereby avoiding modeling each of the many particles explicitly. Instead, 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 sound probability is constant at each time step. This gives rise to an exponential probability weighing time between events. The four signals had different features in terms of duration, amplitude, temporal evolution, and spectrum (see Fig. 3).

Since both males and females were involved in the experiment, footstep sounds were synthesized in order to avoid any specific cue about the gender of the walker, i.e., trying to simulate a sound which could generally be accepted as genderless. This was achieved by modeling the contribution of a type of shoe which fitted for both males and females, as ascertained in a previous gender recognition experiment (Turchet and Serafin 2013).

Three different sound reproduction techniques were considered: monophonic (mono, M), stereophonic (stereo panning, S), and binaural reproduction (B). In the diotically presented mono condition, the peak level of the sounds was set to 55.4, 57.8, 54.2, and 61.5 dB(A) for snow, gravel, wood, and metal, respectively;[2] these sound levels were taken as reference for the other reproduction conditions (S and B).

The stereo signals were obtained by adding half the mean interaural level difference (ILD) of a KEMAR mannequin (Burkhard and Sachs 1975) at $\pm 5°$ azimuth to the ipsilateral channel and subtracting the same half-ILD from the contralateral channel.[3] The 5° value qualitatively corresponds to the displacement of each foot from the median vertical plane, allowing differentiation of left foot from right foot.

Regarding binaural reproduction, a mixed structural modeling (MSM) approach (Geronazzo et al. 2013) to the construction of HRTFs was used. This approach was preferred over individual HRTF measurement because it simulates a typical application scenario where it is not feasible to individually collect HRTFs (a procedure which strictly requires specific hardware, anechoic spaces, and long collection times) and because of the inherent difficulty in measuring and interpreting HRTF data for low-elevation sources such as our own footsteps. By the MSM approach, we approximate the influence of the listener's body on the incoming sounds through a pair of non-individual HRTFs [either generic or selected from the CIPIC database (Geronazzo et al. 2014)] and the optional addition of a spherical torso approximation accounting for shadowing effects on sources coming from below (Algazi et al. 2002). The combination of such choices gave rise to four more reproduction conditions:

1. nonparametric binaural reproduction (B–NP): HRTFs of a KEMAR mannequin;
2. parametric binaural reproduction (B–P): HRTF selection of the best CIPIC subject according to an anthropometry-based distance metric (details follow);
3. nonparametric binaural reproduction with torso (B-NPT): B-NP plus a spherical torso approximation;
4. parametric binaural reproduction with torso (B–PT): B–P plus a spherical torso approximation.

The drawback with non-individual HRTFs such as the KEMAR's is that such peculiar transfer functions will probably never match with the listener's unique anthropometry, and especially his/her outer ear (Spagnol et al. 2013, 2014), resulting in frequent localization errors such as front/back reversals, elevation angle misperception, and inside-the-head localization (Wenzel et al. 1993; Møller et al. 1996). Still, a previous study (Middlebrooks 1999) highlighted the high correlation between the pinna cavity height, i.e., the distance from the superior internal helix border to the intertragic incisure, and an optimal frequency scaling factor aligning spectral HRTF features between subjects and thus minimizing intersubject spectral differences. We used such insight knowledge to guide the selection of the optimal HRTF set in the CIPIC database for a specific subject. Following the CIPIC database anthropometric parameters, the pinna cavity height p_h is given by the sum of d_1 (cavum concha height), d_3 (cymba concha height), and d_4 (fossa height). A simple "best match" of the mean measured p_h between the left and right pinnae detected the best subject for condition B–P (Geronazzo et al. 2013).

Considering the impulsive nature of the footstep sound, one single spatial position for the left and right HRTFs is sufficient. Since no HRTF data for very low elevations are generally available in any public HRTF database because of the difficulty in measuring and interpreting it (Algazi et al. 2002), the lowest-elevation HRTFs were considered in all conditions. These correspond in the CIPIC database to the interaural polar coordinates $(\theta_l, \phi_l) = (-5°, -45°)$ and $(\theta_r, \phi_r) = (5°, -45°)$ for the left and right foot, respectively, where θ denotes azimuth and ϕ denotes elevation.

It has to be recognized that since the used HRTFs were measured at knee height, the elevation impression given to

[2] Such values were chosen according to the results of a previous experiment whose goal was to find the appropriate level of amplitude for those synthesized sounds (Turchet and Serafin 2013). Measurements were conducted by placing the microphone of an SPL meter inside one of the two headphones: Such microphone was inserted in a hole, having its same diameter, created in a piece of hardwood which was subsequently sealed against one of the two headphones. The amplitude peak value of the footstep sound was considered.

[3] The mean ILDs were extracted from the CIPIC HRTF database (Algazi et al. 2001).

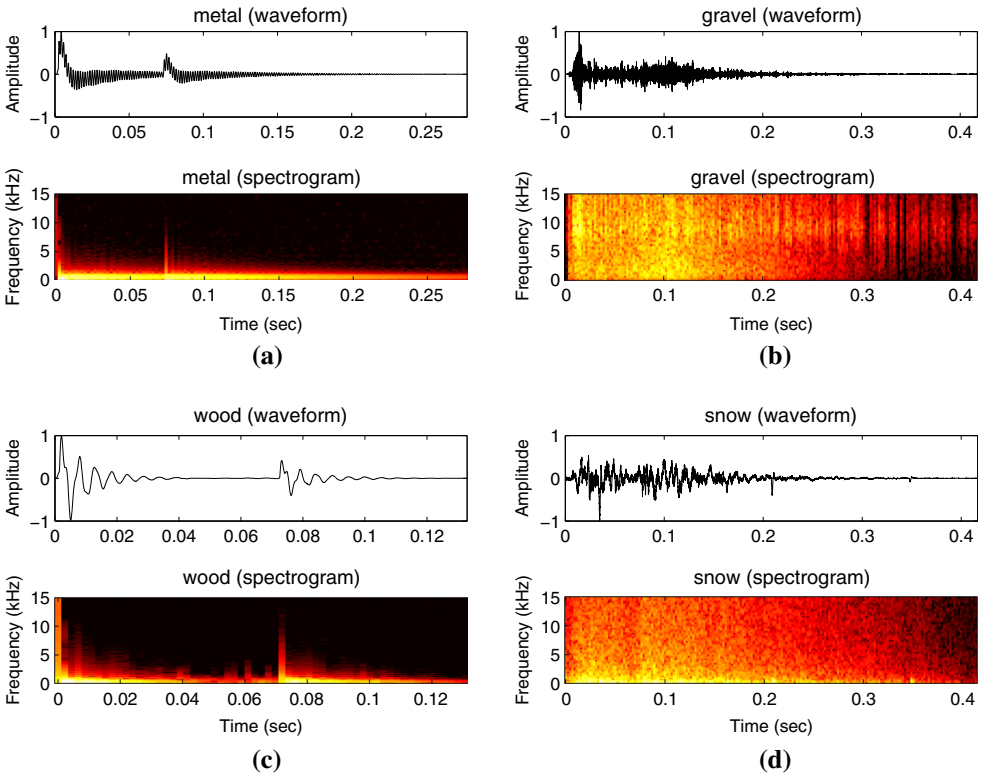

Fig. 3 Typical waveforms and spectrograms of the four simulated materials: **a** metal, **b** gravel, **c** wood, **d** snow

the listener might not be accurate. However, following the simplified geometry of the spherical torso approximation (Algazi et al. 2002), we assumed that the sound wave coming from below travels around the sphere spanning an angle $\theta_{\text{inc}} = 135°$ before reaching the ear(s) at approximately $-45°$ elevation. This approximation was considered in the B-NPT and B-PT conditions, where the theoretical solution for diffraction around a rigid sphere (Strutt 1904) with $\theta_{\text{inc}} = 135°$ was used to design a FIR filter reproducing its magnitude behavior. The only independent variable of the spherical model, i.e., the sphere radius, was adapted to the maximum circumference t_c of the subject's torso.

In order to guarantee the best localization accuracy possible, even to the detriment of perceived realism, no reverberation was applied to the sound stimuli. The combination of the six rendering techniques and the four surface materials gave rise to 24 stimuli, each repeated twice for a total of 48 trials. Trials were randomized across participants.

2.4 Procedure

Participants were first subjected to a short anthropometric measurement session where parameters p_h and t_c were acquired. Then, each subject wore the pair of shoes and a belt which allowed the wires from shoes and headphones to be fixed to the user's back and to then be directed to the Arduino board. In addition, wires were attached to the subject's trousers with Velcro tape and secured to the waist. The wires were long enough (5 m) to allow free motion in the experimental space. The experiment was conducted in a laboratory whose floor was covered with a thin carpet in order to mask the footstep sounds resulting from the interaction of sandals with the floor. Such a masking was further enhanced by the use of the closed headphone set, in addition to the softness of the sandals' sole.

Participants, who were never informed about which material was simulated at each trial, were instructed to walk freely inside the walking area and listen to the headphone-provided footstep sounds as long as they wanted before concluding the trial. At the end of each trial, participants were provided with a printed questionnaire and required to fill the following items:

- Q1 Indicate in Fig. 4 the circlet corresponding to the direction where the sound came from;
- Q2 Evaluate the degree of realism of the sounds you have produced;
- Q3 Evaluate to what extent your way of walking seems natural to you;
- Q4 Evaluate to what extent you feel confused or disoriented while walking.

The choice of a graphical self-report instead of a verbal report is due to avoiding cognitive factors when having to represent the elevation of a sound source. Similar reporting

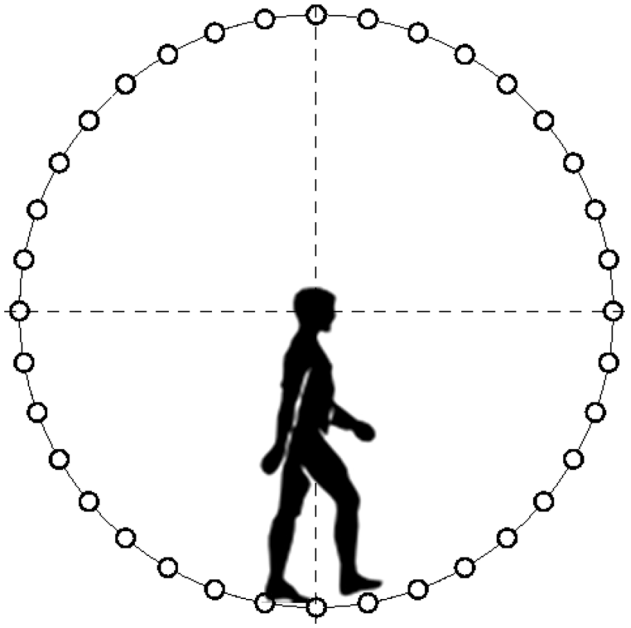

Fig. 4 Figure for questionnaire item Q1

methods for source elevation are commonly found in the literature of 3D auditory localization (Begault et al. 2001; Hwang et al. 2008).

The circlets in Fig. 4 indicate sound location relative to the listener and are 10° equally spaced because of the high localization uncertainty in the median vertical plane (Blauert 1983). Notice that, although the subject moves, the use of headphones guarantees that the virtual location of the footstep sound never changes with respect to the subject himself. Questions Q2, Q3, and Q4 were instead evaluated on a visual analog scale (VAS) [0 = not at all, 10 = very much]. Such questions were motivated by the necessity of having additional information concerning the subjective experience of interacting with the provided virtual world. Specifically, they were chosen because the realism of the provided sounds, the naturalness of the walking experience, and the sense of confusion or disorientation while walking are factors related to the sense of presence (Slater et al. 2009; Turchet 2015).

Before performing the task, subjects were presented with six practice trials, one for each rendering technique, in order to become familiar with the system. To this purpose, the forest underbrush material was chosen [delivered at 53.5 dB(A)]. This material was not among those involved in the experiment.

2.5 Results and discussion

Data corresponding to questionnaire item Q1 were first analyzed with respect to scores corresponding to the circlets placed in the front and back half-circumferences (FHC and BHC) in Fig. 4 (i.e., the points in which the sound was perceived as coming from the front and from the back, respectively). Such an analysis was performed in order to verify the presence of a preference for localization of the sound at the front or at the back. The number of scores in FHC and BHC was counted for each technique and each material separately and subsequently analyzed by means of an exact binomial test. This statistical analysis revealed that in all cases the difference between the counts in FHC and BHC was not significant. Localization scores in the two half-circumferences (negative scores $[-18, 0]$ anticlockwise in the BHC and positive scores [0, 18] anticlockwise in the FHC, where 0 is the lowest point in Fig. 4) were then subjected to a Friedman test for each of the six levels of rendering technique. No significant main effect was found. As a consequence, the localization scores corresponding to BHC were normalized in absolute value and added to those in FHC for further analyses. The resulting data were subjected to three Friedman tests, for rendering technique, for material, and for rendering technique for each material. Only the main effect of material was significant, $\chi^2(3) = 27.7$, $p < 0.001$.

As illustrated in the top-left panel of Fig. 5, the post hoc analysis, performed by using the Wilcoxon–Nemenyi–McDonald–Thompson test, revealed that localization scores for the four materials were all significantly different except between the gravel and snow conditions. In particular, localization scores for the snow and gravel conditions were both significantly lower (i.e., toward the feet of the human silhouette in Fig. 4) than the metal and wood conditions. For the sake of brevity, in the remainder of the paper results of the post hoc tests (all conducted by means of the Wilcoxon–Nemenyi–McDonald–Thompson procedure) are reported in the figures.

Figure 5 also shows the evaluations expressed as VAS scores for questions Q2 (realism), Q3 (naturalness), and Q4 (disorientation) considering the data grouped by material. The three questionnaire items were subjected to a Friedman test for rendering technique and material. Concerning Q2, the main effect of rendering technique was nonsignificant, while the main effect of material was $\chi^2(3) = 23.3$, $p < 0.001$. The post hoc test paralleled that of localization scores, indicating that realism scores were significantly different among all conditions and in ascending order for the metal, wood, gravel, and snow conditions. As regards Q3 and Q4, a significant main effect was again found only for material (Q3: $\chi^2(3) = 15.4$, $p < 0.01$, Q4: $\chi^2(3) = 11.1$, $p < 0.05$). The results of the post hoc test are illustrated in Fig. 5.

In addition, linear mixed-effects model analyses were performed in order to search for correlations between each localization score (in absolute value) and each VAS

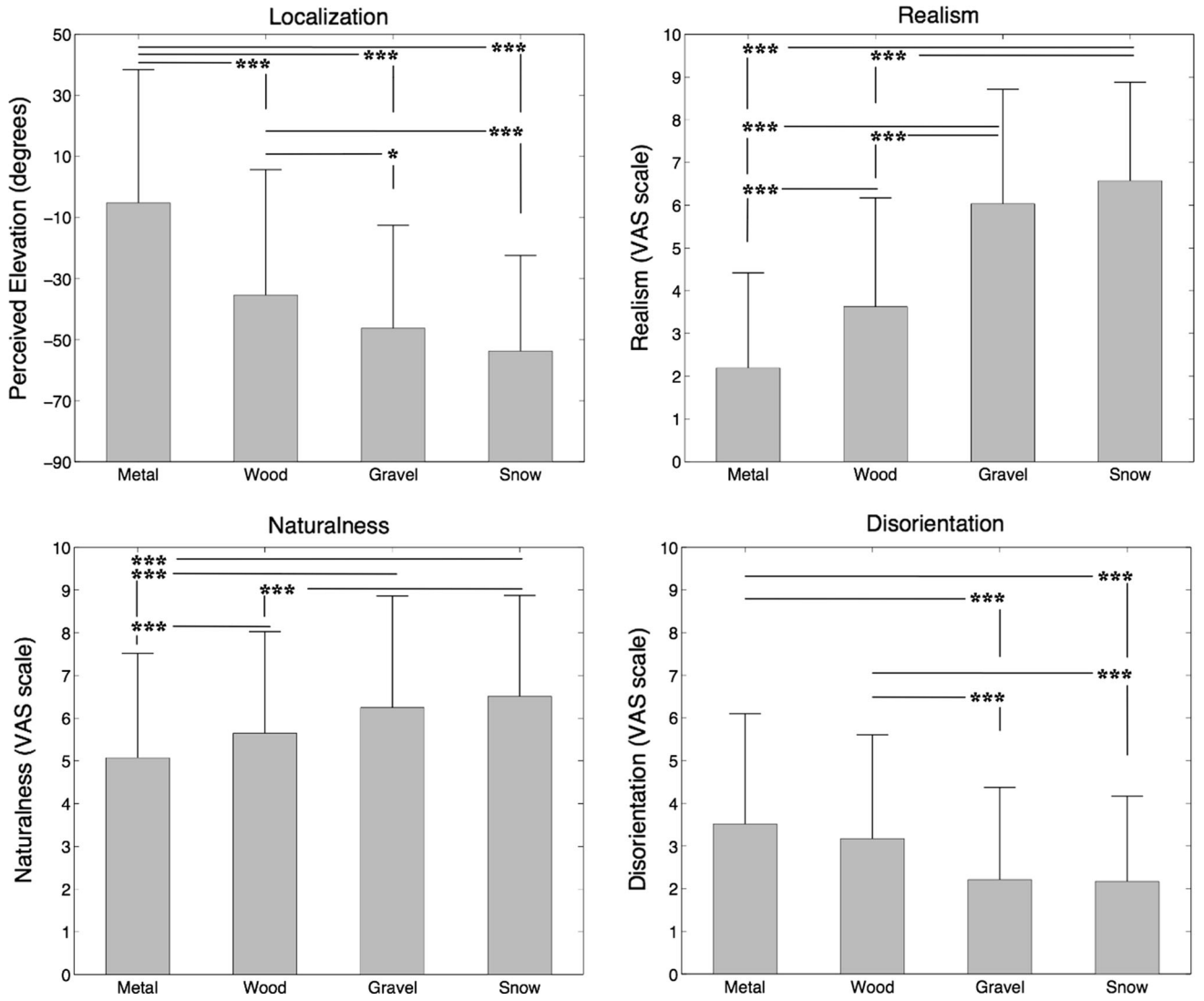

Fig. 5 Results of experiment 1: graphical representation of the mean and standard deviation for questionnaire items Q1 (*top-left*), Q2 (*top-right*), Q3 (*bottom-left*), and Q4 (*bottom-right*). $*p \leq 0.05$; $***p \leq 0.001$

evaluation expressed for Q2, Q3, and Q4. Such analyses revealed that the localization scores were linearly related to perceived realism ($\beta = -7.23$, $t(563) = -13.39$, $p < 0.001$), naturalness ($\beta = -5.1$, $t(563) = -5.93$, $p < 0.001$), and disorientation ($\beta = 5.58$, $t(563) = 6.72$, $p < 0.001$).

The four questionnaire items were then subjected to a Wilcoxon signed-rank test having two levels of surface typology (solid and aggregate). In all cases, a significant main effect was found, showing that localization and disorientation scores were higher for the solid typology compared to the aggregate one ($Z = 10.178$, $p < 0.001$ and $Z = 7.691$, $p < 0.001$ respectively), and realism and naturalness scores were lower for the solid typology compared to the aggregate one ($Z = -15.519$, $p < 0.001$ and $Z = -6.163$, $p < 0.001$ respectively).

No significant differences among the six rendering techniques were found. This is in accordance with our initial hypothesis for solid surfaces, whose associated sounds do not have enough energy at high frequencies to enable vertical localization mechanisms (Hebrank and Wright 1974). As Fig. 3 shows, the frequency content of solid footstep sounds (wood and metal) only overshoots the 4–5 kHz threshold that enables vertical localization by the pinna in very short temporal windows. For footstep sounds in particular, the presence of high-frequency energy is needed to trigger not only pinna-related elevation cues (i.e., frequency notches), but also torso-related ones (i.e., shadowing effects).

However, binaural techniques were all unexpectedly found to be ineffective also for aggregate surfaces,

independently of the degree of customization. Instead, results showed that materials belonging to the aggregate surface typology were always localized significantly lower than the solid ones. Therefore, taken together these results suggest that surface typology has an influence on the localization judgments and that such an influence is strong enough to mask differences between the involved rendering techniques.

Coherently, significant differences were also found between evaluations of aggregate and solid surfaces as far as the perceived realism of the simulations is concerned, as well as the naturalness of the walk and the degree of confusion or disorientation. As illustrated in Fig. 5, those judgments scaled monotonically with the localization scores, and regression analyses proved the presence of linear correlations in all cases.

3 Experiment 2

In order to test the strength of the surface typology effect in localization perception and to confirm the results of the first experiment concerning the absence of differences in localization judgments between the rendering techniques, a second experiment was designed. Specifically, the directionality of footstep sounds was studied in the presence of sonically simulated virtual environments, i.e., adding a soundscape.

The role of contextual information, sonically provided as soundscape, on the perception of footstep sounds was studied by Turchet et al. (2010). Soundscapes sonically simulated either the environment typically associated with the surface material synthesized (i.e., coherently) or with a totally different one (i.e., incoherently). Results showed that adding a coherent soundscape significantly improved both recognition of surface materials and realism evaluations when compared to both footstep sounds alone and with footstep sounds with an accompanying incoherent soundscape.

In our experiment, adding auditory information concurrent to the footstep sounds might decrease the accuracy of their localization, and such a decrement could be greater when incoherent soundscapes are provided compared to the case in which coherent ones are involved. However, if the effect is still present in such conditions this would mean that the effect is strong and that its causes might not only be due to the auditory channel per se but should be searched in the multimodal perceptual mechanisms involved in locomotion.

3.1 Participants

Twelve participants, six males and six females, aged between 19 and 26 ($M = 22.66$, SD $= 2.49$), not one of whom was involved in the previous experiment, took part in the experiment. All participants reported normal hearing and no impairment in locomotion.

3.2 Stimuli and procedure

The same apparatus was used as in the previous experiment. In addition to footstep sounds, the soundscapes of the following four environments were used: a courtyard of a farm during summer; a ski slope; a house interior; and a submarine. Such ad hoc built soundscapes were the same adopted by Turchet et al. (2010) and were chosen in order to coherently fit with the synthesized footstep sounds (gravel, snow, wood, and metal, respectively). When incoherently provided, they were coupled with metal, wood, snow, and gravel, respectively. The used soundscapes were designed so as to provide a clear indication of the designed environments after the first few seconds.

The RMS amplitudes of the soundscapes were set to 54.1, 67.2, 62.7, and 63 dB(A) for the house, the submarine, the courtyard, and the ski slope, respectively. Such values were again chosen according to the results of Turchet and Serafin (2013), whose goal was to find the appropriate sound level for those soundscapes in the presence of synthesized footstep sounds set to the amplitudes indicated in Sect. 2.3.

The experimental protocol was analogous to that of the first experiment. The training phase consisted of presenting the footstep sounds of forest underbrush alone, with a coherent soundscape corresponding to a forest, and with an incoherent soundscape corresponding to a beach seaside in summer. Both the material and the two soundscapes were not among those involved in the experiment.

Footstep sounds were rendered using the M and B-PT techniques only. This choice was made in order to check whether the delivery method affects the quality of the results as far as the aggregate surfaces are concerned in the presence of an accompanying soundscape. Results were expected to confirm those of the first experiment, i.e., no significant differences between M and B-PT. The combination of the two rendering techniques, the four surface materials, and the three soundscape conditions (coherent, incoherent, and no soundscape) gave rise to 24 stimuli, each repeated twice for a total of 48 trials. Trials were randomized across subjects.

3.3 Results and discussion

Results of the second experiment are illustrated in Fig. 6. Localization scores were analyzed by means of a Friedman test for stimulus type (footstep sounds alone, with coherent soundscape, with incoherent soundscape), rendering technique, material, and for rendering technique for each

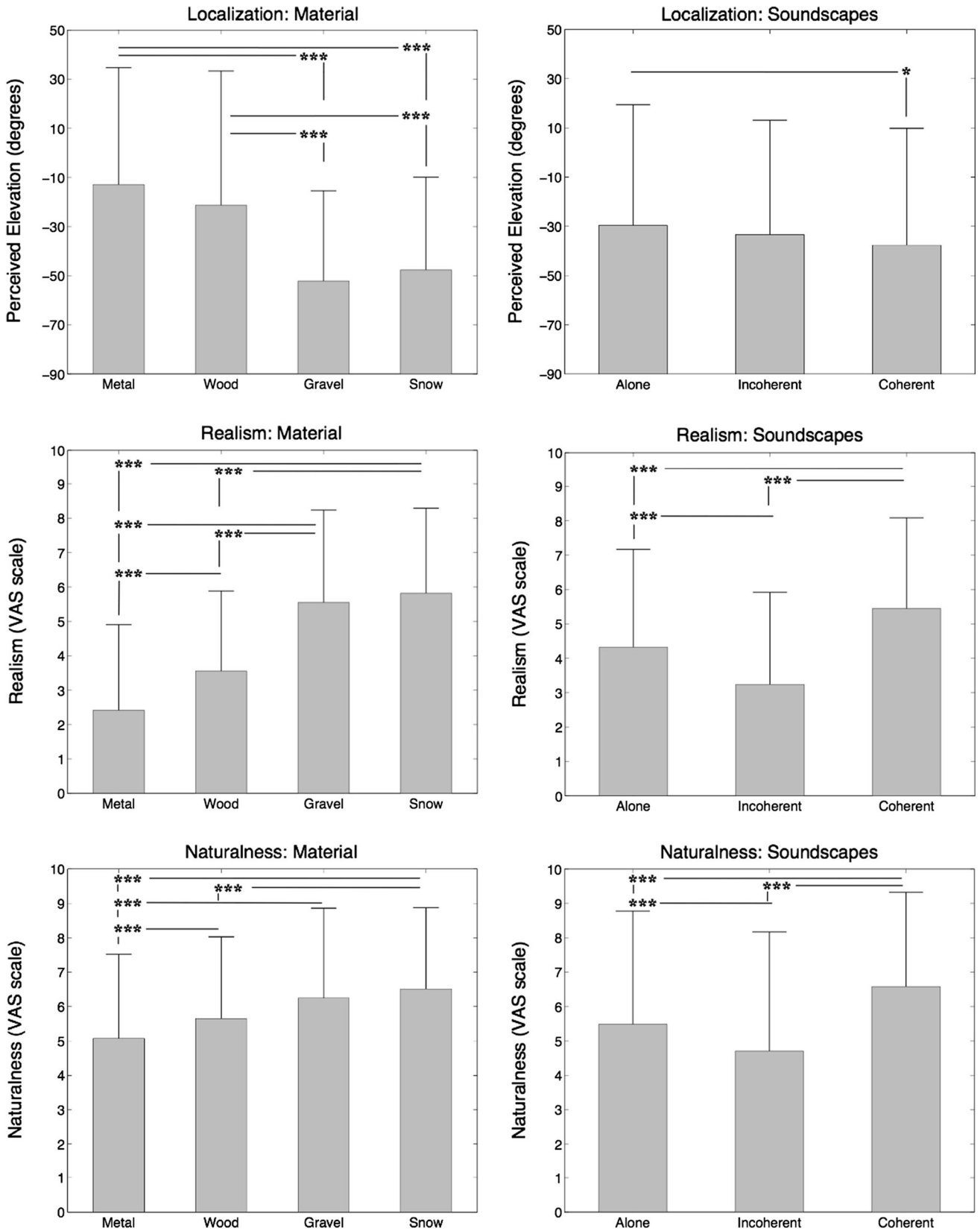

Fig. 6 Results of experiment 2: graphical representation of the mean and standard deviation for questionnaire items Q1, Q2, Q3, and Q4 analyzed by material (*left*) and by type of stimulus (*right*). $^{**}p \leq 0.01$; $^{***}p \leq 0.001$

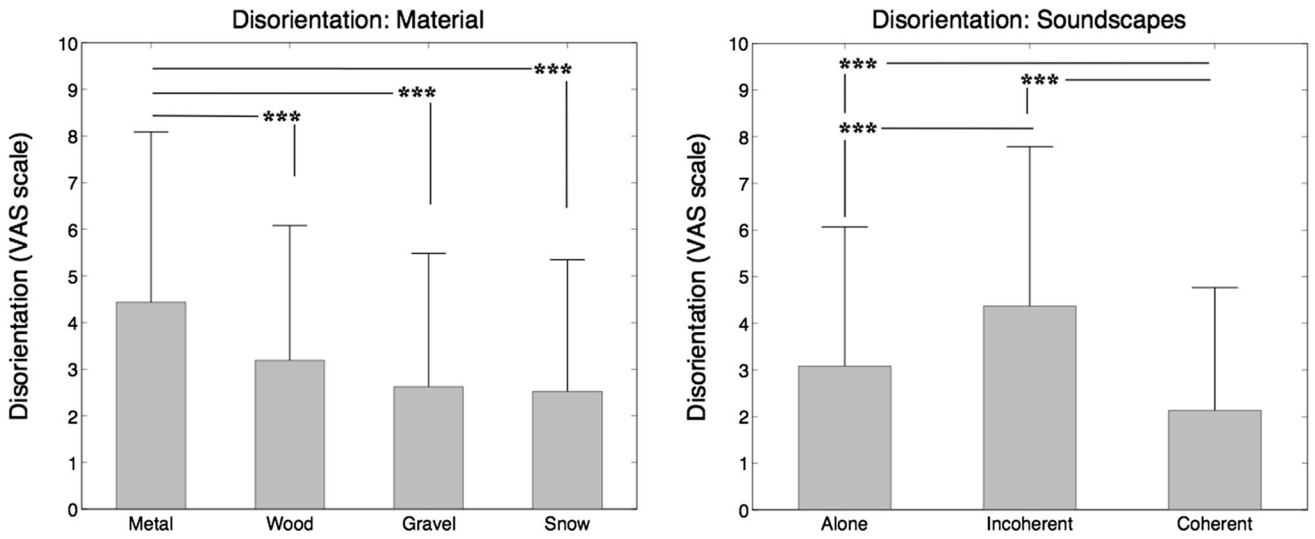

Fig. 6 continued

material. A significant main effect was found for material ($\chi^2(3) = 23.319$, $p < 0.001$). The post hoc analysis revealed that localization scores were significantly lower for both the snow and gravel conditions when compared to both the metal and wood conditions.

The evaluations of Q2, Q3, and Q4 were subjected to the same analyses. As regards Q2, a significant main effect was found for material ($\chi^2(3) = 23.5$, $p < 0.001$) and stimulus type ($\chi^2(2) = 11.166$, $p < 0.01$). The first post hoc test indicated that realism scores were all significantly different except between the gravel and snow conditions; in particular, scores for the snow and gravel conditions were both significantly higher than the metal and wood conditions. The second post hoc test indicated that realism scores were significantly higher for coherent soundscapes when compared to footstep sounds alone and incoherent soundscapes and lower for incoherent soundscapes compared to footstep sounds alone.

Concerning Q3, a significant main effect was found for material ($\chi^2(3) = 23.5$, $p < 0.001$) and stimulus type ($\chi^2(2) = 11.166$, $p < 0.01$). The first post hoc test indicated that naturalness scores were significantly lower for the metal condition when compared to all of the other conditions. The second post hoc test indicated that naturalness scores were significantly higher for coherent soundscapes when compared to footstep sounds alone and incoherent soundscapes and lower for incoherent soundscapes compared to footstep sounds alone. Regarding Q4, a significant main effect was found for material ($\chi^2(3) = 11.533$, $p < 0.01$) and stimulus type ($\chi^2(2) = 11.555$, $p < 0.01$). The results of the two post hoc test were analogous to those of Q3.

The analyses performed with linear mixed-effects models revealed that localization scores (in absolute value) were linearly related to perceived realism ($\beta = -5.97$, $t(563) = -9.73$, $p < 0.001$), naturalness ($\beta = -3.08$, $t(563) = -3.9$, $p < 0.001$), and disorientation ($\beta = 3.02$, $t(563) = 4.1$, $p < 0.001$).

The four questionnaire items were then subjected to a Wilcoxon signed-rank test having two levels of surface typology (solid and aggregate). In all cases, a significant main effect was found, showing that localization and disorientation scores were higher for the solid typology compared to the aggregate one ($Z = 8.974$, $p < 0.001$ and $Z = 5.421$, $p < 0.001$ respectively), and realism and naturalness scores were lower for the solid typology compared to the aggregate one ($Z = -11.41$, $p < 0.001$ and $Z = -6.5479$, $p < 0.001$ respectively).

The results of this second experiment confirm, as expected, the prevalence of the information related to surface typology over the spatial rendering technique as far as perceived localization is concerned. Independently of the surface typology, localization scores were only slightly affected by the presence of a soundscape (precisely by the coherent soundscapes provided compared to the case of footstep sounds alone). Analogously to the findings of the previous experiment, they were linearly related to judgments of realism, naturalness, and disorientation. These results, therefore, indicate that localization of footstep sounds is affected by the simulated surface typology and that this effect is roughly independent of the presence of a soundscape. Concerning the perceived realism of footstep sounds, an influence of the presence of contextual information was noticed: Footstep sounds accompanied by a coherent soundscape were judged significantly more realistic than when provided alone or with an incoherent soundscape. These findings confirm the results reported by Turchet et al. (2010). The results of both the first and

second experiments thus suggest that the influence of surface typology on localization judgments is a robust effect, since it is independent of the used rendering technique and of the presence of contextual information.

4 Experiment 3

The set of surface materials involved in the previous two experiments was relatively small. Only four synthesized materials were used, and no comparison against recordings of real footstep sounds was conducted. Another critical point arising from the first two experiments is that at signal-level aggregate sounds are significantly longer in time and significantly richer in high-frequency content than solid sounds; hence, the found effect could be merely dependent on temporal or spectral factors.

From all these considerations, a third experiment was designed with the goal of (1) replicating the results of the first two experiments using a larger palette of surface materials; (2) testing the effectiveness of synthesized footsteps sounds compared to recorded samples; and more importantly (3) assessing whether the found effect could be due to signal-level features of the involved sound stimulus.

4.1 Participants

Twelve participants, three males and nine females, aged between 19 and 39 ($M = 25.75$, SD $= 6.09$), all of whom were not involved in the previous experiments, took part in this experiment. All participants reported normal hearing and no impairment in locomotion.

4.2 Stimuli and procedure

The same apparatus was used as in the first two experiments. Both recordings of real and synthesized footstep sounds were used, for a total of 21 surface materials (9 solid, 10 aggregate, and 2 control conditions). In particular, the solid materials were wood, concrete, and metal all provided as real and synthesized [54.2, 56.3 and 61.5 dB(A), respectively] sounds. Moreover, three sounds were created by coupling the synthesized materials with a reverberation tail corresponding to a room of size $9 \times 9 \times 2.5\text{m}$ ($T_{60} = 0.505\,\text{s}$). Concerning the aggregate materials, the following surfaces were used (all provided as real and synthesized): snow, gravel, dry leaves, dirt pebbles, and forest underbrush [55.4, 57.8, 54.4, 53.5, 53.5 dB(A), respectively]. The same amplitude for the corresponding real and synthesized materials was adopted and set according to the amplitude indicated in previous research (Turchet and Serafin 2013). The recordings of real surfaces were the same as those used in a previous recognition experiment (Nordahl et al. 2010).

The recordings of real footstep sounds were used to increase the sonic palette and to search for possible differences with the synthesized sounds in the four questionnaire items. Analogously, the addition of reverberation to synthesized solid surfaces was used in order to verify possible differences in participants' evaluations compared to synthesized solid surfaces without reverberation: Indeed, the duration of the reverberated stimuli lasted for a time long enough to cover the average duration of real footsteps, i.e., the whole temporal duration of the haptic stimulus, as opposed to the drier unreverbed sounds.

Moreover, two control conditions were considered. They consisted of white noise bursts, lasting 80 and 420 ms, respectively, both provided at 56 dB(A). The two durations were set to the minimum and maximum duration of the involved solid and aggregate surface sounds, respectively, while amplitudes were set to the average amplitude of all sounds. These control conditions were chosen to verify possible localization biases due to the stimulus' duration or frequency content. As a matter of fact, one of the salient differences between footstep sounds on aggregate and solid surfaces is the duration, which is longer for the first compared to the second. Furthermore, noise bursts have more high-frequency content than aggregate surface sounds; hence, if frequency content were responsible for the localization bias then the noise bursts would be localized even lower.

Since the previous experiments revealed no significant differences between the techniques used for sound delivery, only one technique, M, was used. Each of the 21 stimuli was repeated twice for a total of 42 trials. Trials were randomized across subjects. The procedure was identical to that of the first two experiments, anthropometric measurements excluded. The training phase consisted of presenting recordings of both real and synthesized footstep sounds on sand delivered at 51.9 dB(A). These stimuli were not among those involved in the experiment.

4.3 Results and discussion

Figure 7 shows the results of the third experiment. Localization scores were analyzed by means of a Friedman test for the three levels of surface typology (control, solid, aggregate), yielding a significant main effect ($\chi^2(2) = 15.5$, $p < 0.001$). The post hoc comparisons indicated that the localization scores were significantly higher for the control condition when compared to solid and aggregate conditions and significantly higher for the solid condition when compared to the aggregate condition.

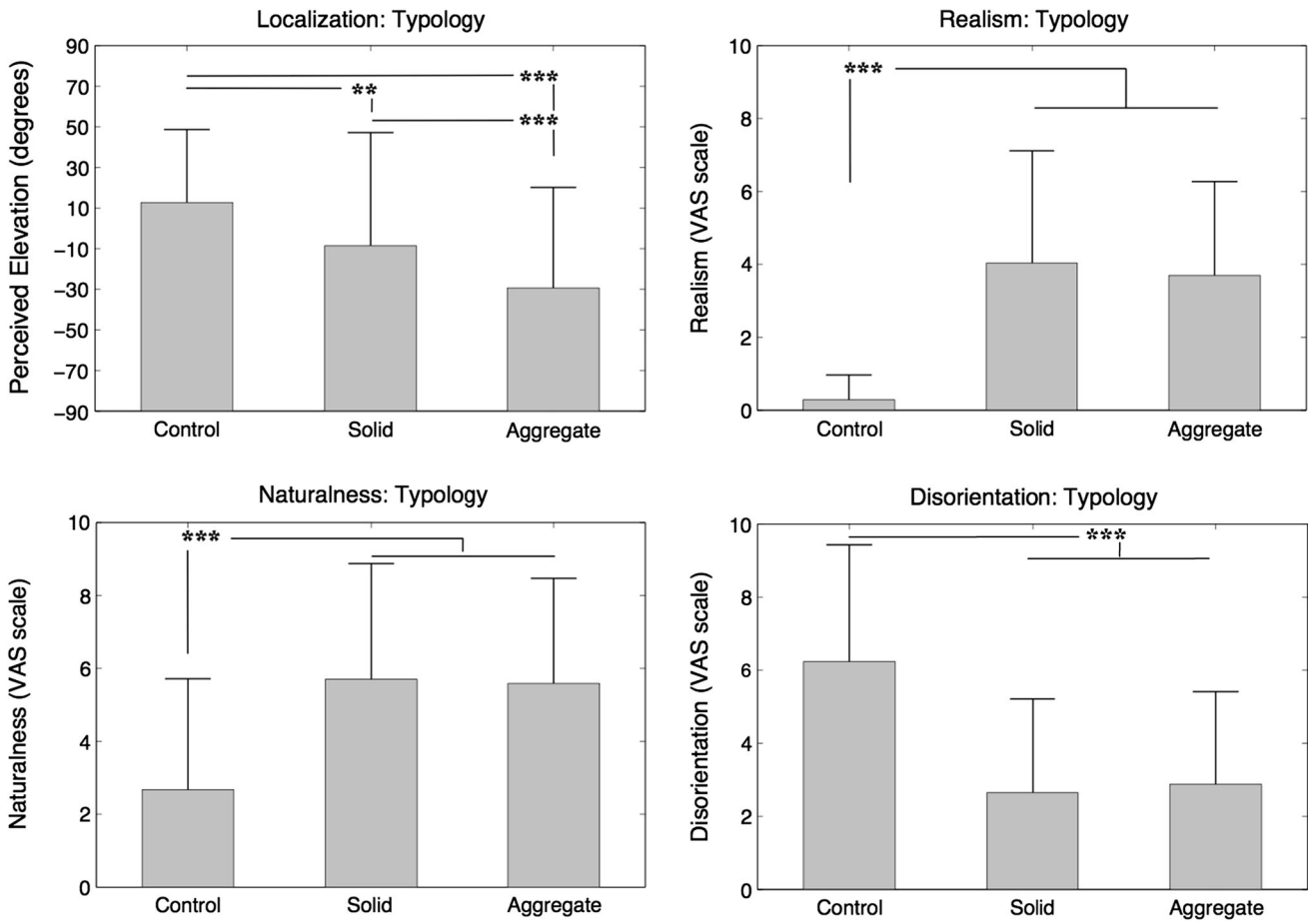

Fig. 7 Results of experiment 3: graphical representation of the mean and standard deviation for questionnaire items Q1, Q2, Q3, and Q4 analyzed by surface typology. $^{**}p \leq 0.01$; $p \leq 0.001$

A Wilcoxon signed-rank test showed no significant differences between localization scores of the synthesized surfaces and the real ones. Similarly, the same test showed no significant differences between localization scores of the synthesized solid surfaces with and without reverberation. Also, no significant differences between localization scores of the two noise bursts were found.

The evaluations of Q2, Q3, and Q4 were subjected to the same analyses. The main effect of surface typology was $\chi^2(2) = 18.666$, $p < 0.001$ for Q2, $\chi^2(2) = 15.166$, $p < 0.001$ for Q3, and $\chi^2(2) = 10.34$, $p < 0.01$ for Q4. The post hoc test indicated that realism and naturalness (disorientation) scores were significantly lower (higher) for the control condition when compared to solid and aggregate conditions, while no significant differences were found either between the synthesized surfaces and the real ones or the synthesized solid surfaces with and without reverberation. A further Wilcoxon signed-rank test was conducted on the four questionnaire items to compare the two control conditions. In none of the analyses, statistical significance was noticed.

The analyses performed with linear mixed-effects models revealed that localization scores (in absolute value) were linearly related to perceived realism ($\beta = -6.34, t(563) = -8.55$, $p < 0.001$), naturalness ($\beta = -4.77, t(563) = -5.4$, $p < 0.001$), and disorientation ($\beta = 4.17, t(563) = 4.94$, $p < 0.001$).

Taken together, results of the third experiment confirm that footstep sounds on aggregate surfaces are localized nearer to the feet than those on solid surfaces. Furthermore, both the noise bursts were localized in positions higher than those corresponding to the real and synthesized solid surfaces, and their localization scores did not differ significantly. Last but not least, no significant localization difference was found between solid surfaces with and without reverberation. Therefore, these findings exclude any explanation of the cause of the found effect due to the duration or frequency content of the sound stimulus.

Contrary to the previous two experiments, realism, naturalness, and disorientation scores were not significantly different for the solid and aggregate surface typologies, while as expected control conditions were judged as the

least realistic. Furthermore, similar ratings were given for the real and synthesized sounds for all the questionnaire items; this suggests the success of the synthesis algorithms in mimicking real footsteps sounds. Analogously, in each of the four questionnaire items no significant difference was found for the synthesized solid surfaces with and without reverberation. This finding parallels the corresponding localization results.

5 General discussion

The main result common to the three experiments is that solid surfaces are localized significantly farther from the walker's feet than aggregate ones independently of rendering technique, presence or absence of contextual information, duration and frequency content of the sound stimulus. Such an effect could be explained by the presence of a semantic conflict between the haptic and auditory sensory channels, coupled with the hypothesis that the auditory system uses the information coming from the haptic channel to enhance sensitivity in the localization of sounds apparently coming from the walker's feet.

Such a hypothesis is inspired to the findings reported by Thomas and Shiffrar (2010) who argued that the visual system could make use of auditory cues during visual analysis of human action (in their case, footsteps) when there is a meaningful match between the auditory and visual cues. In our study, the source of the auditory and haptic (i.e., the foot-shoe contact while walking) stimuli was not unique, and therefore, the two sensory channels received conflicting information. Still, our interpretation is supported by the evidence that audiotactile interactions can happen independently of spatial coincidence in the region close to the head (see Kitagawa and Spence 2006 for a review) and parallels the findings on how information presented on one sensory modality can influence information processing in another sensory modality [e.g., the ventriloquism illusion (Howard and Templeton 1966) and the "parchment-skin" illusion (Jousmäki and Hari 1998)]. However, it is interesting to notice that since the apparent location of the presented footstep sounds is not particularly biased toward the source of the synchronous tactile stimulation (i.e., the feet), the phenomenon of tactile capture of audition (Caclin et al. 2002) does not happen.

Our hypothesis is further fostered by the findings reported by Laurienti et al. (2004) that highlighted how the semantic content of a multisensory stimulus plays a critical role in determining how it is processed by the nervous system, and by the results recently reported by Turchet and Serafin (2014). That study presented a set of experiments whose goal was to investigate subjects ability to match pairs of synthetic auditory stimuli (created with the same engine used in the present work) and haptic stimuli [delivered through haptic shoes (Turchet et al. 2010)]. The involved stimuli were both semantically congruent (e.g., wood delivered at both auditory and haptic level) and incongruent (e.g., snow delivered at haptic level and metal simultaneously delivered at auditory level) and presented in both active (i.e., while walking) and passive (i.e., while sitting on a chair) sensorymotor activity. Results showed that in the active condition pairs of stimuli consisting of an auditory aggregate material and a haptic solid material were not judged less semantically congruent than pairs of solid materials, as well as of aggregate materials. Conversely, aggregate–solid pairs were judged, with statistical significance, to be less semantically congruent than solid–solid pairs in the passive condition. The cause for this result was attributed to technological limitations. Indeed, although the impact sound produced by hard sole shoes with a solid surface was realistically rendered, haptic stimuli induced by the actuators were not effective in masking the haptic sensation due to the softness of the sandals' sole and the presence of a carpeted floor. This is also the case of the current study, in which haptic shoes simulating solid surfaces were not even used.

Therefore, we hypothesize that the haptic sensation arising when walking with sandals over a floor covered with carpet is more semantically incongruent with the simultaneous presentation of an impact sound between a hard sole and a solid surface than with the simultaneous presentation of a footstep sound on an aggregate surface. From this, it follows that the different localization ratings reported in the present study could be attributable to the different levels of semantic congruence between auditory and haptic information: The lower the semantic congruence, the greater the distance of the sound source from the feet.

Besides the described incongruence between auditory and tactile information, there are two more sources of conflicting multisensory information that could have contributed to the found effect. The first concerns the role of vision. This hypothesis is supported by different studies on the ventriloquism effect (Howard and Templeton 1966) that showed an influence of visual cues on auditory localization. During all trials subjects could see the carpeted surface which they were walking upon as well as the whole laboratory space. These visuals could have created an expectation of sound that corresponds to walking on a carpet in an indoor environment, violated in the presence of the delivered auditory feedback. According to this hypothesis, the greater the discrepancy between the heard sound and the expected sound the higher the perceived localization. Although this hypothesis was not systematically investigated in the reported experiments, our current results do not support it. First, incongruence is always

present for all stimuli, as none of them simulates a carpeted surface. Second, aggregate surfaces should produce the highest auditory–visual conflict (because they are associated with outdoor environments), but according to our results these sounds produce the lowest localization scores and the highest degrees of realism.

The second source of conflicting multisensory information regards the role of proprioception. In fact, previous research highlighted cross-modal effects between audition and proprioception while walking on a solid surface with sandals and listening to the sound of an aggregate material, such as an alteration of the perceived softness of the walked-upon surface and the induction of a sense of sinking (Turchet et al. 2013, 2015). Again, the higher is the conflict between auditory and proprioceptive information, the higher the source should be perceived. However, also in this case our current results do not support this hypothesis, as auditory information of solid surfaces is more congruent with the proprioceptive information given by the solid surface of the laboratory compared to that of the aggregate ones. More precisely, metal, wood, gravel, and snow can be ordered by increasing compliance, whereas in our results perceived elevation increases with decreasing compliance.

It is undoubted that different levels of congruence between the involved sensory information (including the contextual information provided as soundscape) produce different levels of presence, as the realism, naturalness, and disorientation scores demonstrate. Consequently, despite our results not supporting visual or proprioceptive effects, the possibility that localization of different surface typologies depends on a combination of the listed incongruences cannot be completely ruled out. In particular, in order to confirm the dominance of auditory–haptic semantic congruence over the above-mentioned conflicts, an experiment could be conducted where subjects wear shoes with a solid sole while walking on an uncarpeted surface. Our hypothesis would predict lower localization scores for solid surfaces compared to the less semantically congruent aggregate surfaces.

In addition, it is worthwhile to notice that the present study involved auditory stimuli both valid and not valid from the ecological point of view. In the presence of non-ecological stimuli (i.e., noise), the location of the sound source was rated higher than the corresponding congruent and incongruent ecologically valid stimuli. This is a further indication that when the association between the information arriving to ears and feet is not meaningful, interaction between the two sensory channels produces percepts which are not reliable. On a separate note, realism, naturalness, and disorientation scores were found to be unaffected by semantic congruence but were linearly correlated with the localization scores in all experiments. This suggests the importance of using interactive footstep sounds that are perceivable as realistic and capable to induce a high sense of naturalness during walking, as well as not to create a sense of disorientation. In short, the ecological validity of the auditory feedback is a relevant aspect in the design of locomotion-based interfaces.

6 Conclusions

These findings have interesting applicative as well as theoretical implications. In terms of designing audio-haptic locomotion interfaces for virtual reality contexts, care should be taken to provide users with feedback fully valid from the ecological point of view, and capable of producing a meaningful association between the two sensory modalities. Our results coupled with the interpretation of previous works suggest that the type of shoe plays a relevant role in the meaningfulness of the association between simulations of auditory and haptic stimuli. This aspect has received scarce attention from designers of synthetic footstep sounds and vibrotactile feedback. Furthermore, our findings suggest that the use of spatial sound reproduction techniques (through headphones) is less relevant than the meaningfulness of bimodal associations.

In practical terms, two are the main implications to the design of locomotion interfaces for virtual reality. The first is that the technology for sound reproduction can be simplified by omitting the simulation of spatial effects. The second is that semantic congruence between auditory and tactile stimuli should be ensured in order to avoid bias in the localization of self-generated footstep sounds when provided through headphones. For this purpose, the tactile shoes presented in Turchet et al. (2010) could be used. Nevertheless, results of Turchet and Serafin (2014) suggest that by means of that technology, which involves soft sole shoes, the haptic rendering of solid surfaces is not as effective as that of aggregates. To cope with this limitation, our results would suggest to wear shoes with hard sole and a non-carpeted solid surface when solid surfaces are delivered at auditory level.

On the other hand, understanding how different perceptual and cognitive factors influence localization of sounds produced by self-generated actions fosters our theoretical understanding of human multimodal perception and cue integration, a field that receives growing research interest. In particular, our results contribute to the development of a theoretical framework of the perceptual mechanisms involved in sonically simulated foot–floor interactions mediated by locomotion interfaces. Ultimately, future research will allow investigation of how audio-haptic interactions in walking contribute to the internal multisensory representation of the body.

Acknowledgments The work of the first author was supported by the Danish Council for Independent Research, Grant No. 12-131985.

References

Algazi VR, Duda RO, Duraiswami R, Gumerov NA, Tang Z (2002) Approximating the head-related transfer function using simple geometric models of the head and torso. J Acoust Soc Am 112(5):2053–2064

Algazi VR, Duda RO, Thompson DM (2002) The use of head-and-torso models for improved spatial sound synthesis. In: Proceedings of 113th convention audio engineering society, Los Angeles, pp 1–18

Algazi VR, Duda RO, Thompson DM, Avendano C (2001) The CIPIC HRTF database. In: Proceedings of IEEE workshop on applications signal processing, audio and acoustic. New Paltz, New York, pp 1–4

Avanzini F, Rocchesso D (2001) Modeling collision sounds: non-linear contact force. In: Proceedings of digital audio effects conference, pp 61–66

Begault DR, Wenzel EM, Anderson MR (2001) Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J Audio Eng Soc 49(10):904–916

Blauert J (1983) Spatial hearing: the psychophysics of human sound localization. MIT Press, Cambridge

Boren BB, Geronazzo M, Majdak P, Choueiri E (2014) PHOnA: a public dataset of measured headphone transfer functions. In: Proceedings of 137th audio engineering society convention

Burkhard MD, Sachs RM (1975) Anthropometric manikin for acoustic research. J Acoust Soc Am 58(1):214–222

Caclin A, Soto-Faraco S, Kingstone A, Spence C (2002) Tactile"capture" of audition. Percept Psychophys 64(4):616–630

Cheng CI, Wakefield GH (2001) Introduction to head-related transfer functions (HRTFs): representations of HRTFs in time, frequency, and space. J Audio Eng Soc 49(4):231–249

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

Geronazzo M, Spagnol S, Avanzini F (2013) Mixed structural modeling of head-related transfer functions for customized binaural audio delivery. In Proceedings of 18th international conference on digital signal processing (DSP 2013). Santorini, Greece

Geronazzo M, Spagnol S, Avanzini F (2013) A modular framework for the analysis and synthesis of head-related transfer functions. In: Proceedings of 134th audio engineering society convention, Rome, Italy

Geronazzo M, Spagnol S, Bedin A, Avanzini F (2014) Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP 2014), Firenze, Italy, pp 4496–4500

Giordano B, Visell Y, Yao HY, Hayward V, Cooperstock J, McAdams S (2012) Identification of walked-upon materials in auditory, kinesthetic, haptic and audio-haptic conditions. J Acoust Soc Am 131:4002–4012

Hebrank J, Wright D (1974) Spectral cues used in the localization of sound sources on the median plane. J Acoust Soc Am 56(6):1829–1834

Howard IP, Templeton WB (1966) Human spatial orientation. Wiley, New York

Hunt KH, Crossley FRE (1975) Coefficient of restitution interpreted as damping in vibroimpact. ASME J Appl Mech 42(2):440–445

Hwang S, Park Y, Park Y (2008) Modeling and customization of head-related impulse responses based on general basis functions in time domain. Acta Acust United Acust 94(6):965–980

Jousmäki V, Hari R (1998) Parchment-skin illusion: sound-biased touch. Curr Biol 8(6):R190–R191

Kitagawa N, Spence C (2006) Audiotactile multisensory interactions in human information processing. Jpn Psychol Res 48(3):158–173

Kobayashi Y, Osaka R, Hara T, Fujimoto H (2008) How accurately people can discriminate the differences of floor materials with various elasticities. IEEE Trans Neural Rehab Syst Eng 16(1):99–105

Laurienti P, Kraft R, Maldjian J, Burdette J, Wallace M (2004) Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 158(4):405–414

Middlebrooks JC (1999) Individual differences in external-ear transfer functions reduced by scaling in frequency. J Acoust Soc Am 106(3):1480–1492

Møller H, Sørensen MF, Jensen CB, Hammershøi D (1996) Binaural technique: Do we need individual recordings? J Audio Eng Soc 44(6):451–469

Nordahl R, Berrezag A, Dimitrov S, Turchet L, Hayward V, Serafin S (2010) Preliminary experiment combining virtual reality haptic shoes and audio synthesis. In: Haptics: generating and perceiving tangible sensations, lecture notes in computer science, Springer, Berlin, vol 6192, pp 123–129

Nordahl R, Serafin S, Turchet L (2010) Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proceedings of the IEEE virtual reality conference. IEEE Press, pp 147–153

Papetti S, Civolani M, Fontana F (2011) Rhythm'n'shoes: a wearable foot tapping interface with audio-tactile feedback. In: Proceedings of the international conference on new interfaces for musical expression, pp 473–476

Papetti S, Fontana F, Civolani M, Berrezag A, Hayward V (2010) Audio-tactile display of ground properties using interactive shoes. In: Haptic and audio interaction design, Lecture notes in computer science, Springer, Berlin, vol 6306, pp 117–128

Serafin S, Turchet L, Nordahl R, Dimitrov S, Berrezag A, Hayward V (2010) Identification of virtual grounds using virtual reality haptic shoes and sound synthesis. In: Proceedings of eurohaptics symposium on haptic and audio-visual stimuli: enhancing experiences and interaction, pp 61–70

Slater M, Lotto B, Arnold MM, Sanchez-Vives MV (2009) How we experience immersive virtual environments: the concept of presence and its measurement. Anuario de Psicologia 40(2):193–210

Spagnol S, Geronazzo M, Avanzini F (2013) On the relation between pinna reflection patterns and head-related transfer function features. IEEE Trans Audio Speech Lang Process 21(3):508–519

Spagnol S, Geronazzo M, Rocchesso D, Avanzini F (2014) Synthetic individual binaural audio delivery by pinna image processing. Int J Pervasive Comput Commun 10(3):239–254

Stein B, Meredith M (1993) The merging of the senses. MIT Press, Cambridge

Steinicke F, Visell Y, Campos J, Lécuyer A (2013) Human walking in virtual environments: perception, technology, and applications. Springer, Berlin

Strutt JW (1904) On the acoustic shadow of a sphere. Philos Trans R Soc Lond 203:87–110

Thomas JP, Shiffrar M (2010) I can see you better if i can hear you coming: action-consistent sounds facilitate the visual detection of human gait. J Vis 10(12):14. doi:10.1167/10.12.14

Turchet L (2015) Designing presence for real locomotion in immersive virtual environments: an affordance-based experiential approach. Virtual Real (accepted)

Turchet L (2015) Footstep sounds synthesis: design, implementation, and evaluation of foot-floor interactions, surface materials, shoe types, and walkers' features. Appl Acoust (in press)

Turchet L, Camponogara I, Cesari P (2015) Interactive footstep sounds modulate the perceptual-motor aftereffect of treadmill walking. Exp Brain Res 233:205–214

Turchet L, Nordahl R, Berrezag A, Dimitrov S, Hayward V, Serafin S (2010) Audio-haptic physically based simulation of walking on different grounds. In: Proceedings of IEEE international workshop on multimedia signal processing, IEEE Press, pp 269–273

Turchet L, Serafin S (2011) A preliminary study on sound delivery methods for footstep sounds. In: Proceedings of digital audio effects conference, pp 53–58

Turchet L, Serafin S (2013) Investigating the amplitude of interactive footstep sounds and soundscape reproduction. Appl Acoust 74(4):566–574

Turchet L, Serafin S (2014) Semantic congruence in audio-haptic simulation of footsteps. Appl Acoust 75(1):59–66

Turchet L, Serafin S, Cesari P (2013) Walking pace affected by interactive sounds simulating stepping on different terrains. ACM Trans Appl Percept 10(4):23:1–23:14

Turchet L, Serafin S, Nordahl R (2010) Examining the role of context in the recognition of walking sounds. In: Proceedings of sound and music computing conference

Visell Y, Cooperstock J, Giordano B, Franinovic K, Law A, McAdams S, Jathal K, Fontana F (2008) A vibrotactile device for display of virtual ground materials in walking. Lect Notes Comput Sci 5024:420–426

Visell Y, Fontana F, Giordano B, Nordahl R, Serafin S, Bresin R (2009) Sound design and perception in walking interactions. Int J Hum Comput Stud 67(11):947–959

Vliegen J, Van Opstal AJ (2004) The influence of duration and level on human sound localization. J Acoust Soc Am 115(4):1705–1713

Wenzel EM, Arruda M, Kistler DJ, Wightman FL (1993) Localization using nonindividualized head-related transfer functions. J Acoust Soc Am 94(1):111–123

Zanotto D, Turchet L, Boggs E, Agrawal S (2014) Solesound: Towards a novel portable system for audio-tactile underfoot feedback. In: Proceedings of the 5th IEEE international conference on biomedical robotics and biomechatronics, pp 193–198