Effects of Interactive Sonification on Emotionally Expressive Walking Styles

Luca Turchet and Roberto Bresin

Abstract—This paper describes two experiments conducted to investigate the role of sonically simulated ground materials in modulating both production and recognition of walks performed with emotional intentions. The results of the first experiment showed that the involved auditory feedbacks affected the pattern of emotional walking in different ways, although such an influence manifested itself in more than one direction. The results of the second experiment showed the absence of an influence of the sound conditions on the recognition of the emotions from acoustic information alone. Similar results were found in both experiments for musically-trained and untrained participants. Our results suggest that tempo and sound level are two acoustical features important in both production and recognition of emotions in walking. In addition, the similarities of the presented results with those reported in the music performance domain, as well as the absence of an influence of musical expertise lend support to the "motor origin hypothesis of emotional expression in music" according to which a motor origin for the expression of emotions is common in all those domains of human activity that result in the generation of an acoustical signal.

Index Terms—Interactive sonification, emotions, walking

1 Introduction

Walking is one of the most typical everyday activities linking body motion and sound production. Lately several research efforts have been devoted to the interactive sonification [1] of walking [2], i.e., the study of the human interaction with a system that transforms data concerning foot-floor interactions into sounds. In particular, the interest in investigating how auditory feedback can influence the patterns of walking is growing in several fields including entertainment, virtual and augmented reality, sports, therapy, diagnosis, and rehabilitation.

For example, an approach used in motor therapy consists of providing a clicking sound in response to each step and to ask patients to keep a rhythmic auditory cue by controlling their gait pattern. This technique was proved to be effective in improving walking abilities in patients affected by multiple sclerosis [3] and cerebral palsy [4]. A different approach based on music played at a speed adjusted to the walker's strides revealed to improve the gaits of hemiparetic stroke patients [5].

In non-clinical context, studies have been conducted on understanding the role of interactive auditory feedback in modulating the pattern of locomotion. In a recent one, a system composed by a footstep sound synthesis engine and augmented shoes with pressure sensors has been used in an uncontrolled outdoor environment [6]. In this study it was ascertained that locomotion is significantly affected when walkers are interactively provided with sounds simulating steps on a terrain different from that they are trampling on. In particular, there was a scaling effect from higher to lower material compliance such that individuals walked faster when the simulated sound resembled wood, than with gravel and snow. The rationale for these results was attributed to three possible plausible explanations: an audio-foot haptic semantic incongruence, an audio-foot haptic temporal conflict, or an adjustment to the perceived sonically simulated surface material.

In a similar vein, the influence of the interactive auditory feedback on the walking style has been also recently investigated in a pilot experiment conducted in the context of emotional walking [7]. Subjects were asked to walk in an indoor environment with four different emotional intentions (happy, sad, aggressive, tender) while listening to their own footstep sounds produced both on real ground materials (wood panels and linoleum) and by an interactive sonification actuated by sensorized-shoes (muddy ground and iced snow). Results, although not statistically significant, provided indications that walking patterns can be influenced by the sound of the ground independently from the emotional intention of the person.

In related research [8], [9], it has also been shown that when walking with different emotional intentions humans make variations of timing and sound level in the same way as found in expressive music performance [10], [11]. For example, it has been found that music performances communicating happiness and happy walking styles were both characterized by a faster tempo/pace and louder sound level relative to a neutral style, while performances and walking patterns communicating sadness were characterized by slower tempo/pace and softer sound level. Such results lent support to the "motor origin hypothesis of

- *L. Turchet is with the Department of Media Technology, Aalborg University Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen. E-mail: tur@create.aau.dk.*
- *R. Bresin is with the School of Computer Science and Communication, KTH Royal Institute of Technology, Lindstedtsvägen 3, SE-11428 Stockholm. E-mail: roberto@kth.se.*

Manuscript received 12 July 2014; revised 9 Mar. 2015; accepted 22 Mar. 2015. Date of publication 25 Mar. 2015; date of current version 3 June 2015.
Recommended for acceptance by G. N. Yannakakis.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2015.2416724

emotional expression in music" formulated by Giordano et al., according to which musicians and listeners make use of general movement knowledge when expressing and recognizing emotions in music: *"emotions are expressed in similar ways in music and vocal performance domains because they are both the result of a motor activity that expresses emotions in coherent ways independent of the particular effector"* [8], [9].

To date, scarce attention has been devoted to the relationship between interactive sonification of walking actions and emotions. Typically, research on emotion production and identification in auditory contexts involves musical stimuli [10], [12]. Ecologic sounds have been rarely utilized in emotion research involving interactive contexts. Therefore, a relevant research question concerns the understanding of the role of ecologic sounds interactively provided, such as footstep sounds, in altering walking patterns in emotional contexts. The findings reported in [6], [7], [8], and [9], motivated us to continue a research to explore the relationship between interactive sounds and movements in the context of the production and identification of emotionally expressive walking sounds. Another open question in auditory research on emotions is whether musical expertise affects the production and identification of expressive walking styles in interactive contexts involving ecologic sounds.

In this paper two experiments are presented. In a first experiment (production experiment) we tested whether by manipulating the auditory feedback delivered to a person walking with different emotional intentions would influence his/her locomotion style. Specifically, we were interested in assessing whether there are synthetic footstep sounds that make people walk with different styles (i.e., faster or slower, stepping harder or softer) both dependently or independently from their emotional intention. Indeed, it is in principle possible that some interactive ecological sounds, such as those of footsteps, can enhance, decrease, or not affect at all the style of emotionally expressive walking. To achieve this goal we designed an experiment using an interactive sonification actuated by a non intrusive shoe-independent system for providing walkers with auditory feedback valid from the ecological point of view [13], [14], [15]. This system could generate walking sounds of different surfaces that were triggered in real-time by the walker's footsteps. Our hypothesis was that, within the same emotion, walkers would change their walking strategy depending on the sound feedback that corresponded to a simulated surface material. Such a hypothesis was motivated by the following considerations.

Firstly, the findings reported in [6] indicated that aggregate-compliant materials, such as snow and gravel, are capable to affect walkers' pace by lowering their speed, while the locomotion is performed on a non compliant surface (which is the case of the present study). Secondly, the results presented in [7] suggested that walkers used a faster pace when the sound of the walking surface was characterized by a higher spectral centroid, and a slower pace when the spectral centroid was low. Also, solid surfaces led to more aggressive walking patterns while aggregate ones to more tender and sad walking styles. Thirdly, studies on expressive music performance showed that articulation is a parameter linked to emotions. Staccato articulation (i.e., when notes are played with shortened duration and are separated from the next note by a silent interval) is associated to emotions with high level of activity (e.g., happiness, aggressiveness) while legato articulation (i.e., when the notes are played smoothly and connected to each other, producing an acoustic overlap) to emotions with low level of activity (e.g., sadness, tenderness) [10], [12]. In another study by Bresin and Dahl [16] it has been shown that timing patterns in staccato articulation in piano performance correspond to those in running, while legato articulation corresponds to pace used in walking. This suggests that staccato articulation is associated to fast tempo in walking, while legato to slow tempo in walking. In the case of footstep sounds, staccato and legato can be related to the degree of distinction of heel and toe. For solid surfaces impacted with hard sole shoes, heel and toe sub-events are more discernible (i.e., staccato) compared to aggregate surfaces (i.e., legato), since in the latter case the sound produced by the heel is prolonged so to mix with the sound corresponding to the toe. In a different vein, in music performance research it has been ascertained that tones with high spectral centroid and/or high pitch ones are often associated with more active emotions (such as happiness and aggressiveness) while their absence and/or low pitched tones are usually associated with low activity emotions (such as sadness and tenderness) [10], [12], [17]. Based on all these considerations, interactive footstep sounds having a predominancy of high frequency components could lead to a more active production of emotions compared to sounds having a predominancy of low frequency components. On the other hand, interactive footstep sounds of aggregate materials could lead to a less active production of emotions compared to solid ones.

Two groups of participants were involved in the first experiment: musically-trained and -untrained. We hypothesized that musicians would have had a better control of their expressive walking compared to non-musicians since musically-trained people are more used to control sound expression through their body in an interactive way [18]. Specifically, we expected that if some sounds adapt better or worse to a certain emotion, musicians would have been more sensitive in expressing through their walking strategies such an influence. Such a hypothesis is furthered fostered by several results in the neuroscience domain that demonstrated how musical training is effective in shaping the brain structure, particularly in brain regions involved in sound perception and production [19], [20].

In a follow up experiment (recognition experiment) we tested to what extent the emotional intentions portrayed by the sounds generated by the participants in the production experiment could be recognized in a listening test, and whether the recognition was moderated by the type of ground material sonically simulated. Recordings of sequences of synthesized footstep sounds produced by walkers in the first experiment were selected according to the criterion of being the most representative for each combination of emotion and ground. As in the first experiment, we used two groups of participants, who did not take part in the first experiment: musicians and non-musicians. The rationale to use these two groups was to assess whether the listeners' expertise yielded different results. As a matter of fact, the importance of the listeners' expertise was risen in [21] for the categorization task of environmental sounds. Since musicians are more exposed to expressive auditory stimuli than non-musicians, they were expected to be more accurate in the discrimination of the emotions.

The final aim of this research was to ascertain whether there are synthetic footstep sounds that affect the expression of acted emotions. Understanding which sounds support or not a specific emotion, would allow to provide guidelines for designers of footstep sounds for virtual reality contexts where the sound feedback associated to foot-floor interactions can influence user behavior. Also, the study is relevant for rehabilitation and therapy applications where the control of the walking style of a patient is a desired goal. For example, it could be that some specific footstep sounds that are effective in eliciting a more active production of emotions, could induce patients to walk with a faster pace depending on the provided auditory feedback: walking speed is often slower in persons with stroke, multiple sclerosis or cerebral palsy and it could be modulated using synthetic footstep sounds as an alternative to current methods involving non interactive auditory stimuli, such as rhythmic auditory stimulation (for a review see [22]), interactive auditory feedback [3], [4], [5] or visual feedback [23].

2 Production Experiment

2.1 Participants

Twenty participants were divided in two groups (n = 10) to perform the experiment. The first group was composed by musicians (4M, 6F), aged between 22 and 46 (mean = 27.3, SD = 7.05). They played different instruments (piano, nyckelharpa, guitar, violin, saxophone, accordion, percussions) or were singers (soprano, mezzo-soprano) with an average musical experience of 14.4 years. The second group was composed by non-musicians (6M, 4F), aged between 23 and 33 (mean = 27.7, SD = 3.12). Participants who belonged to this group never played any musical instrument nor took music lessons. However, they reported to listen to music on average every day. Participants had different nationalities (mostly Europeans). All participants reported neither hearing nor locomotion problems. The average duration of the experiment was about 48 minutes. Participants were compensated for their participation.

2.2 Apparatus

A non intrusive shoe-independent system was developed for the real-time synthesis of footstep sounds (the setup is illustrated in Fig. 1). The system, installed in an acoustically isolated semi-damped room, consisted of a wooden plank (5.40 m long, 90 cm wide and 7 cm tall), under which a set of nine microphones was placed to detect the feet-floor interaction. The microphones (Line Audio Design CM3) were arranged on the floor at an equal distance of 60 cm from each other along the line lying at the half of the plank width. In particular, each microphone was attached to the floor by means of scotch tape, and was lifted of 2 cm from the floor thanks to a layer of foam. The microphones signals were mixed together (through a Yamaha 01V Digital Mixing Console), and subsequently digitalized by means of a sound card (Fireface UFX) connected to a laptop running the footstep sounds synthesizer. The real footstep sounds produced by a walker were detected by the microphones and used to control the temporal evolution of the synthetic footstep sounds according to the real-time synthesis technique developed in previous research [24], [25]. This technique were based on a model where an exciter signal was fed into a resonator. The exciter corresponded to the envelope extracted from the signal captured from the microphones by using the filter described in Section 2.5. The resonator corresponded to physical models for solid object collisions, such as the Impact Model [26], and for particle interactions, such as the Physically Informed Stochastic Model [27]; the former were used for simulating sounds generated while walking on solid surfaces (e.g. wood), the latter for sounds generated when walking on aggregates (e.g., gravel).[1] The synthesized sounds were finally conveyed back to the walker by means of a closed headphone set (Akg K 518 Dj). The headphones' wire connected to the output of the sound card was extended by means of a 3 m long audio cable and attached to the ceiling in order to allow the maximum freedom of movement during the experiment. Even though wired headphones were involved, participants were barely aware about their presence during walking since the wire was light, long enough to move freely, and did not constitute any major constraint for the performance. More importantly, subjects could not hear their own footstep sounds, because they were wearing closed headphones, and the amplitude of the interactively synthesized footstep sounds was high enough to completely mask them.

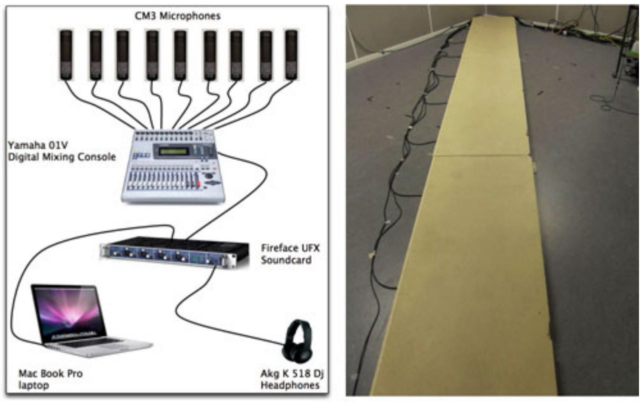

Fig. 1. The block diagram of the microphones-based system (left) and the picture of the wooden plank (right) under which the array of microphones was placed.

Such a system was designed and developed with the goal of satisfying the requirements of fidelity in the accuracy of the foot movements, shoe independence, and freedom of navigation. A solution based on microphones allows for the maximum in dynamics in the detection of the walker's movements and consequently in the production of the corresponding synthesized sounds, compared to other systems previously developed for the same purposes [24]. In addition, placing the microphones under the wooden plank instead of outside it, allows the detection of the sole interaction between the feet and the floor, since this approach avoids the detection of other sound sources (such as the rubbing of pant leg material, the squeak of shoes, or even voices and coughing), which constitute unwanted input

1. Sound excerpts of the sounds simulated throughout the synthesis engine can be found on http://www.ahws-project.net/

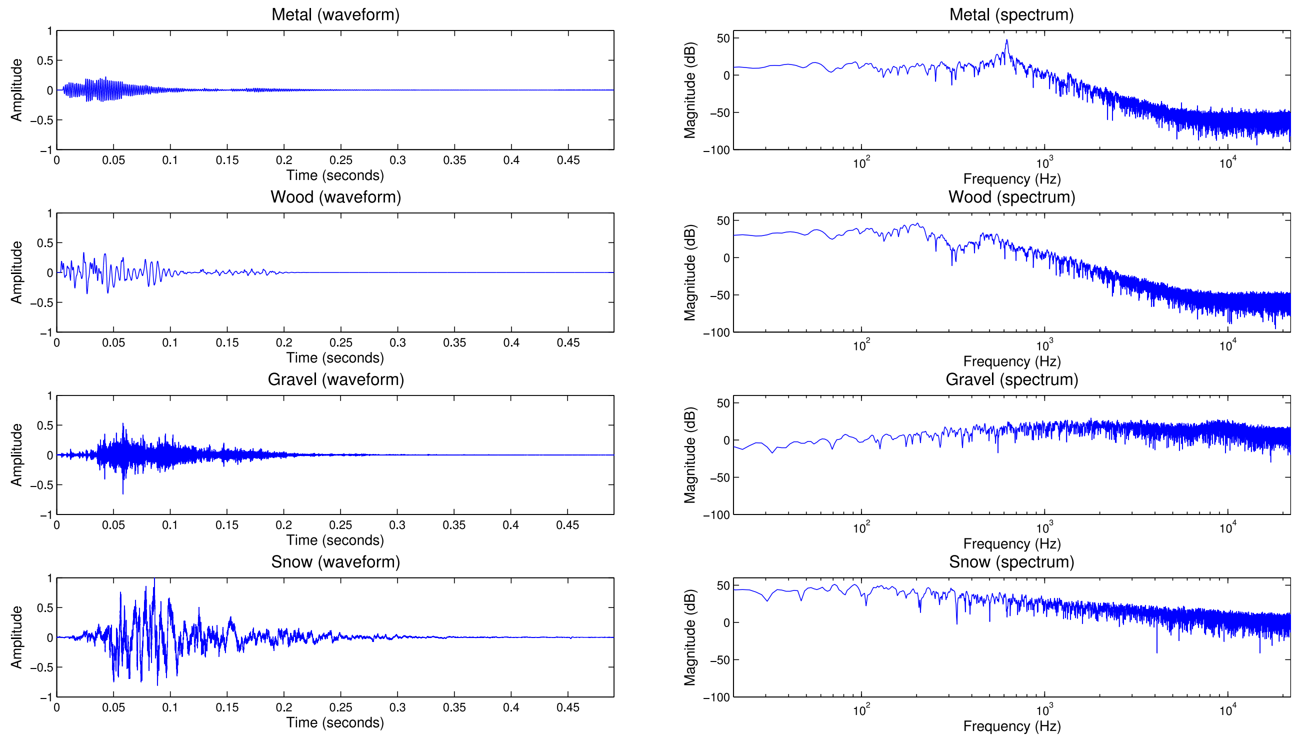

Fig. 2. Typical waveforms (left) and spectra (right) of the four simulated materials. The sounds are synthesized from the same footstep sound generator.

signals for the sound synthesis engine. Moreover, a correct mapping between the sound level of walkers' footsteps and that of the corresponding interactively synthesized sounds, was obtained by assuring an identical detection accuracy in each point of the wooden plank. Furthermore, the developed apparatus allows the walkers to wear their own footwear, contributing to the ecological validity of the interaction [25].

As far as the the synthetic auditory stimuli are concerned, their ecological validity was assessed in previous research [15]. Results of both interactive and non-interactive listening experiments showed that most of the surfaces synthesized using the proposed footstep sound engine were recognized with high accuracy. In particular, they were proven to be correctly classified in the corresponding solid or aggregate surface typology. Similar accuracy was noticed in the recognition of recorded real footstep sounds, which was an indication of the success of the proposed algorithms and their control.

All these aspects allowed to accomplish an ecologically-valid human-system interaction, which was a fundamental prerequisite for the purposes of the experiment, since an interaction poor from the ecological point of view might inhibit both the emotional production and reaction.

2.3 Stimuli

The synthesis engine was set to simulate footsteps on two solid (wood and metal), and two aggregate materials (snow and gravel). Such sound stimuli were chosen for three reasons. First, they presented different features in terms of duration, amplitude, onset type, temporal evolution, and spectrum (see Fig. 2). In more details, the centroids (i.e., the geometric center of the spectrum) were placed at about 380 Hz for wood, 700 Hz for metal, 4,200 Hz for snow, and 9,000 Hz for gravel. Moreover, the two aggregate materials were characterized by a bandwidth much larger than that of the solid surfaces. Second, they were proven to be easy to recognize, as well as to classify in the corresponding solid and aggregate surface typology [15]. Third, they were proven to express different degrees of compliance [6].

The adopted synthesis technique allows to generate sounds whose amplitude is related to the amplitude of the footstep sound captured by the microphones, which in turn is related to the impact force of the foot on the floor. Due to a simulation based on physical models the range of amplitude variation was different for each ground. Sounds were delivered through headphones in the following ranges (in dB (A)), as measured with a sound level meter: [67.78, 86.59] for metal, [74.1, 91.5] for wood, [45.22, 87.26] for gravel, and [52.26, 96.76] for snow.[2]

2.4 Procedure

Participants were asked to wear the headphones and to walk along the wooden plank with five different emotional intentions: happy, aggressive, tender, sad, and neutral. For each intention the ground texture sound was manipulated four times to simulate metal, wood, gravel and snow. Participants were presented with a graphical user interface made by numbered buttons, each labelled with one of the five emotions. They were instructed to press each button

2. Measurements were conducted by placing the microphone of the 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.

according to its numerical order, and to perform the walk with the indicated emotional intention. The five walking styles were repeated twice for each of the synthesized surface materials, and performed in randomized order, for a total of 40 trials per participant. During each trial participants were asked to walk six times along the plank (back and forth three times), for a total walking length of about 30 meters. This number of walks was necessary for limiting the emotional influence of the previous trial, and for assuring a sufficient exposition time to the current auditory stimulus.

For each emotion, participants were instructed to imagine specific situations as follows:

Sad. "You are walking in a cemetery during the funeral of a dear friend"

Happy. "It is a wonderful sunny day, you have won the lottery and you are walking towards the lottery headquarter to get the money"

Tender. "You are walking carrying a three months old baby in your arms"

Aggressive. "You are angry with your neighbor since the loudness of his music does not allow you to sleep, so you are walking towards his flat to ask him for the umpteenth time to lower the volume down"

Neutral. "Walk normally, without any emotional intention"

The emotions happiness, sadness, aggressiveness, and tenderness were chosen mainly because they have been investigated in several studies on emotional expression in music (for an overview see [28]), and because they cover the four quadrants of the two-dimensional Activity-Valence space [29]. Furthermore, these four emotions give a quite comprehensive overview on the use of musical parameters tempo and sound level, that are parameters that can be found and analyzed also in walking [8], [9]. The neutral emotion was used as a control condition. In previous studies on expressive music performance, tempo and sound level in emotionally neutral performances received average values larger than in sad and tender performances, and smaller than in happy and aggressive performances [10].

Participants wore a pair of their own shoes, selected so as to generate relatively loud walking sounds that could be better detected by the microphones, especially during soft walks. Before the beginning of the recording session participants practiced each walking style twice, and tried all the auditory stimuli in order to familiarize with the system. Neither feedback concerning the appropriateness of the adopted walking style nor the information of the type of simulated material were given.

To ascertain the level of individual awareness achieved after the end of the experiment, we asked participants to fill a questionnaire consisting of three parts. The first part had to be completed before seeing the questions of the second part, and consisted of the following question to be rated on seven-point Likert scale ([1 = not at all, 7 = very much]):

Q1. "Rate to what extent the sound influenced the way you were walking during the experiment"

Subsequently, participants were asked to motivate and comment their answer. The second part consisted of the following questions, presented in randomized order, to be rated on a seven-point Likert scale [1 = not at all, 7 = very much]

Q2. "Rate to what extent the type of sound influenced the way you were walking. Type of sound = impact sound (simulating a solid surface) vs granular sound (simulating a non-solid surface)"

Q3. "Rate to what extent the duration of the sound influenced the way you were walking"

Q4. "Rate to what extent the volume of the sound influenced the way you were walking"

Q5. "Rate to what extent the frequency of the sound influenced the way you were walking"

In the third part, participants were asked to recognize the type of sound (solid or aggregate), and to name the simulated surface materials. Subsequently, they were asked to order the four sounds by their ascending influence exerted on the walking style during the experiment. Finally, they were given the possibility to leave an open comment about their experience.

2.5 Acoustical Features Analysis

The footstep sounds produced in each trial were recorded and those corresponding to the fifth walk were retained for the subsequent analysis. In particular, the first two and the last two footstep sounds of that sequence were removed from the analysis, since the initial and final steps of a walk are often atypical (they contain respectively accelerando and rallentando temporal patterns). Two acoustical features were calculated from the collected data: the average time interval between two adjacent heel strikes (in the following called H2H, i.e., "heel-to-heel"), and the average sound level peak (PK) of the recorded footstep sounds. The choice of considering only these two features was motivated by the findings reported in [8], [9]. Indeed, other descriptors, such as the duration of each footstep sound and its spectral content were discarded since in walking it is not possible their strict control.

The calculation of H2H and PK was based on the envelope (e) extracted from the signal (x) captured by the microphones by means of the non-linear low-pass filter proposed in [30]:

$$e(n) = (1 - b(n))|x(n)| + b(n)e(n-1)$$

$$b = \begin{cases} 0.8 & \text{if } |x(n)| > e(n-1) \\ 0.995 & \text{otherwise,} \end{cases}$$

where n and $n-1$ indicate respectively the current and previous sample (sample rate 44,100 Hz) of the discretized variable they refer to, and b_{up} and b_{down} are equal to 0.8 and 0.995 respectively.

An ad-hoc algorithm was implemented in order to identify the portions of the amplitude envelope corresponding to each footstep sound. First of all, only the samples of the envelope signal above a given threshold defining the background noise were considered. The impulsive variations of the envelope signal were analyzed for detecting the starting point of a footstep sound. For each sample above the background noise threshold the first derivative was calculated. If the value of the first derivative was above a given threshold defining the impulsive nature of the variation, the next 20 ms were analyzed in order to find a peak in the signal above a given threshold defining the heel strike. Envelope peaks whose level did not exceed that given threshold, or

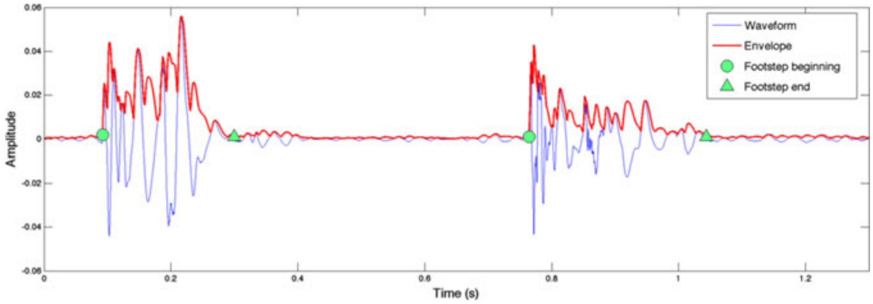

Fig. 3. Identification of the beginning and of the end of two subsequent footstep sounds.

above-threshold peaks lasting less than 20 ms, were disregarded since they had a different origin than the foot impact on the floor. In order to detect the end of the footstep sound, the last value of the signal above the background noise threshold was searched in the range delimited by two thresholds that defined the minimum and the maximum temporal duration for a footstep sound. Finally, the subsequent footstep sound was searched in the portion of the signal after a threshold defining the minimum temporal distance between temporally adjacent footsteps.

Once the footstep sounds were isolated (see Fig. 3) the two acoustical features H2H [ms] and PK [dB] were calculated. The peak level of each footstep was computed taking the maximum of the sound levels computed in overlapping buffers according to the following formula proposed in [31] (using buffers of 20 ms with a hop size of 10 ms):

$$SL_i = 20\,log\frac{\sum_{n=ih+1}^{ih+N}[x(n)w(n)]^2}{\sum_{n=ih+1}^{ih+N} w(n)^2} \quad dB,$$

where i is the buffer index, N the number of samples in a buffer, h the hop size, x is the original signal, n the sample index, and w is a Hanning window with size N.

Based on findings reported in previous research (see Section 1), we expected an influence of the auditory feedback on the performance of each emotional intentions, in terms of either an increase or decrease of the two investigated parameters. Given the spectral properties of the involved sounds metal and gravel were expected to produce a more active performance for each emotion compared to wood and snow. Vice versa, wood and snow were expected to produce for each emotion a less active performance. In addition, we hypothesized that, for each emotion, aggregate materials would have increased H2H, while solid materials would have decreased it. Consequently, snow and gravel were expected to adapt better to sad and tender walking styles, while wood and metal to aggressive and happy ones.

2.6 Results

Fig. 4 illustrates the experimental results for the two investigated parameters and for the two groups of participants. Statistical analysis was performed on the collected data of the two groups by means of repeated measures ANOVA. All post-hoc analyses were performed by using Tukey's procedure (p-value was set at a significant $p < 0.05$). Both the collected H2H and PK were subjected to a three-way analysis of variance having five levels of emotion (happy, aggressive, tender, sad, and neutral), four levels of ground (metal, wood, gravel and snow), and two levels of musical expertise (musicians and non-musicians).

A significant main effect of H2H was found for emotion, $F(4,72) = 39.98$, $p < 0.001$. The pairwise comparison showed that H2H was significantly greater for aggressive compared to neutral ($p < 0.05$), tender ($p < 0.001$), and sad ($p < 0.001$), for happy compared to tender ($p < 0.001$) and sad ($p < 0.001$), and neutral compared to tender ($p < 0.001$) and sad ($p < 0.001$). There was not significant main effect either for factors ground and musical expertise, nor for any interaction effect. Similarly, a significant main effect of PK was found for emotion, $F(4,72) = 4.78$, $p < 0.001$. The pairwise comparison showed that PK was significantly greater for aggressive compared to happy ($p < 0.001$), neutral ($p < 0.001$), tender ($p < 0.001$), and sad ($p < 0.001$), for happy compared to tender ($p < 0.001$) and sad ($p < 0.001$), and neutral compared to tender ($p < 0.001$) and sad ($p < 0.001$). There was not significant main effect either for factors ground and musical expertise, nor for any interaction effect.

From these results it is not possible to conclude that there was a strong influence of the different sounds on the gait patterns produced during the walks performed with an emotional intention. However, it is not possible to conclude that the sound did not have an influence on the participants' gait style, since in this situation the use of a mean blurs the evidence and does not properly represent the underlying phenomena. Indeed, an analysis on the individual performances revealed that on average participants produced, for each emotion, gaits different in terms of H2H and PK depending on the provided sounds. Nevertheless, the effects of the four sounds manifested themselves in more than one direction for each individual. In order to quantify the effect of the sound on the gait patterns for each emotion, the variance for both H2H and PK was calculated by considering all sound conditions together. Results are illustrated in Table 1. As it is possible to notice, the variance of the two parameters were quite substantial. As far as the variation of H2H is concerned, results seem to indicate that emotions were affected differently by the auditory feedback. However, such findings need to be related to the actual speed held by walkers while performing the different emotions. Indeed, great variations of H2H are possible only for slower paces such as those held for sad and tender, while fast walks like those produced for happy and aggressive do not allow variations of H2H in the same range of the slow walks. An analogous consideration can be made for PK. Therefore, it is not possible to conclude that each emotion was affected in a different way by the auditory feedback. Conversely, the entity of the effect of the sounds was

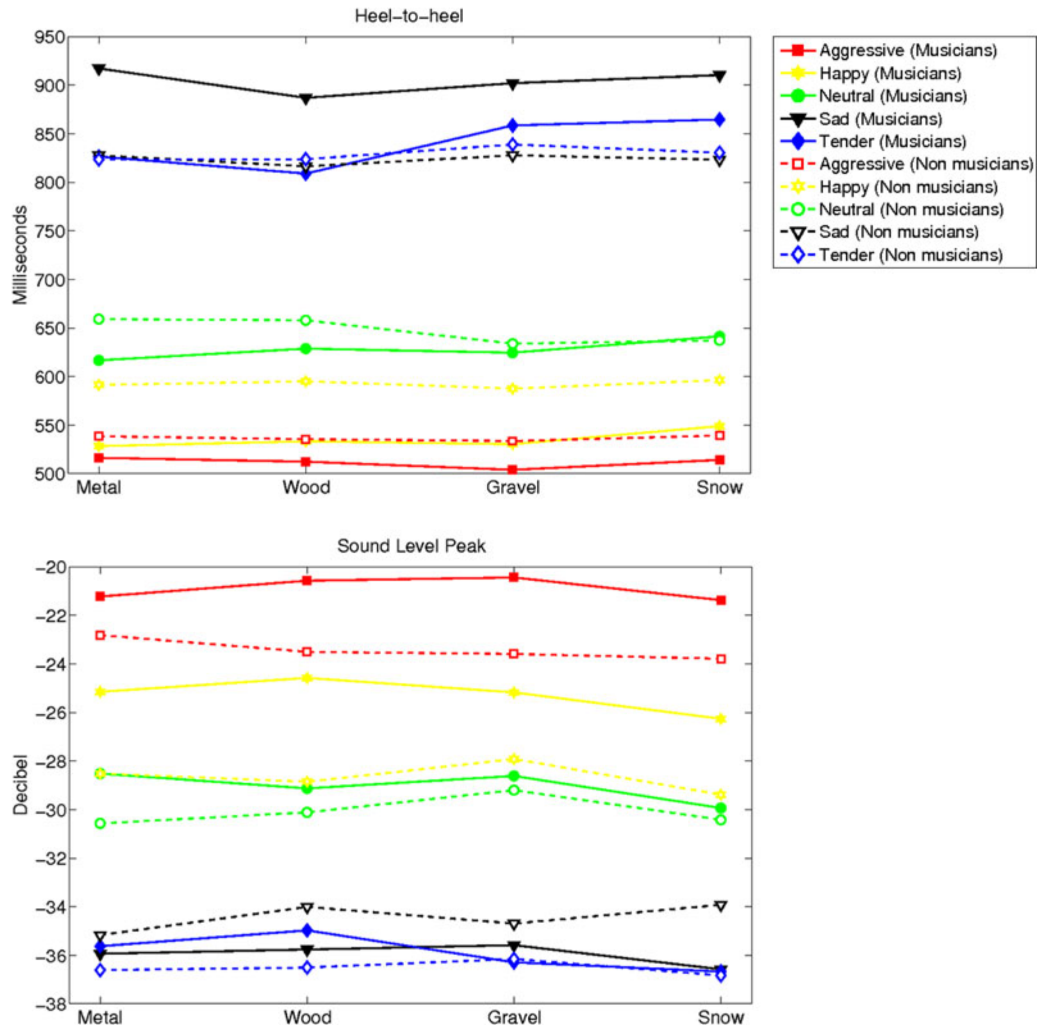

Fig. 4. Results of the production experiment.

basically constant for all the five walking styles. Furthermore, a one-way ANOVA was conducted in order to assess if the differences between the two groups of participants were significant. No significant main effect was found between musicians and not musicians considering both H2H and PK variance.

As concerns the questionnaire, Fig. 5 shows the evaluations expressed by both the groups of participants for each questionnaire item. A Friedman Test was performed for each group of participants separately in order to assess if the differences between the ratings of the questionnaire items Q1, Q2, Q3, Q4, and Q5 were significant, yielding to a non significant main effect. In addition, no significant main effect was found between the two groups as result of a Mann-Whitney-Wilcoxon Test conducted for each questionnaire item.

Interestingly, subjects reported several comments at the end of the walking experience. Those relevant to the influence of auditory feedback on the walking strategies are listed below, divided for each group:

Musicians. "Snow and gravel allowed me to walk better when performing slow paces for sad and tender"; "It was hard to walk happy or tender in presence of snow, while it was hard to walk sad with metal and gravel that are sounds that brought me to walk happy"; "The sound of snow was heavy and it was difficult to walk in a tender way with it. Snow enhanced the aggressive walks. Probably I stepped harder in presence of metal when I was asked to walk aggressive"; "I tried to ignore the sound, but I am aware that sound had an influence on my walking style. It was easier to walk aggressively with hard sounds like snow, and easier to walk tenderly with softer sounds like metal".

Non-Musicians. "It seems to me that snow affected how hard I stepped and wood how fast I walked"; "I tried to ignore the influence of the sound focusing on the rendering of the emotion"; "Snow made me walk more aggressively"; "Snow makes me feel like I am walking hard, metal makes me feel like if I am walking happy or tender"; "Gravel set

TABLE 1
Variance of Both H2H and PK Calculated for Each Emotion

Emotion	H2H (ms)	PK (dB)
Sadness	199	4.2
Tenderness	199	5.2
Neutral	73	3.8
Happiness	62	3.1
Aggressiveness	50	2.7

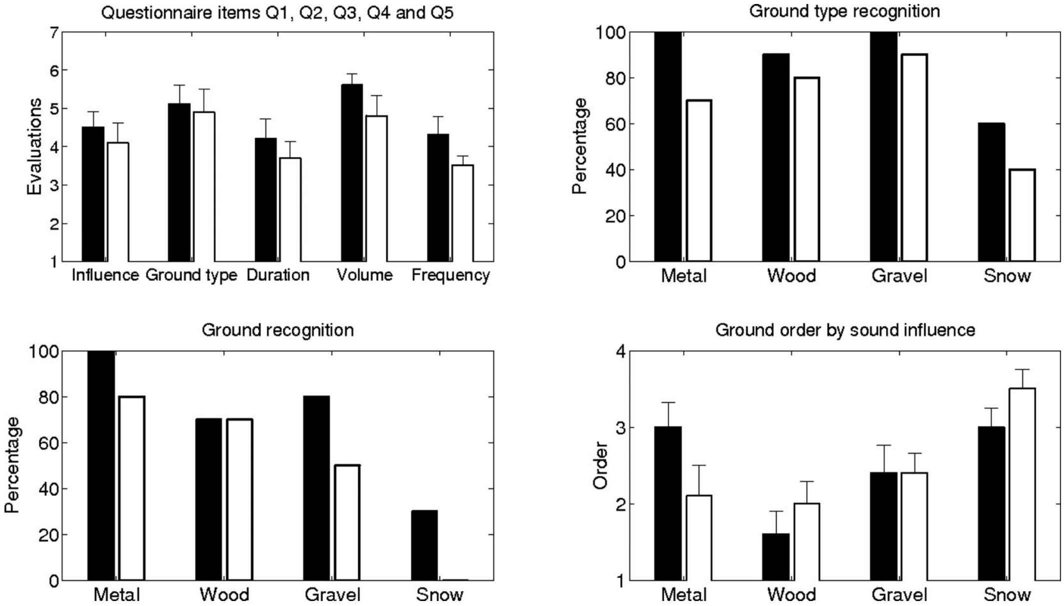

Fig. 5. Graphical representation of the mean and the standard error for participants' answers to questionnaire items Q1, Q2, Q3, Q4 and Q5 (top-left), percentage of recognition of the ground type (top-right), percentage of recognition of the ground material (bottom-left), and order of the grounds by sound influence (bottom-right). In black the ratings of the musicians and in white those of the non-musicians.

me in the happy mood"; "Some of the sound increased the feeling that I was asked to walk as."

2.7 Discussion

The results of the first experiment indicate that subjects walked using different speeds and impact forces for each emotional intention, and that such choices were independent from the particular surface material sonically simulated. This confirms the findings reported in [8], [9] where different emotions were rendered by walkers using different sound level and tempo variations similar to those reported in studies about expressive music performance [11]. In addition, from Fig. 4 it is possible to notice that the neutral emotion is placed in the middle, as also found in a recent study in expressive music performance [10].

Nevertheless, the sound had an influence on the participants' gait style, although the effect of each ground texture was not homogeneous across participants. Indeed, the effect of the sound of a particular ground manifested itself in more than one direction for the same emotion, both in terms of the speed of the walks and of the force with which the ground was hit. The main reason for this behaviour may lie in the fact that participants performed the experiment with different degrees of involvement in the simulations. Indeed, from the open comments reported by participants at the end of the experimental session it emerges that six of them ignored the sound on purpose. These participants motivated this choice with the will to focus on the rendering of emotions in the best possible way. However, the same participants clearly stated that it was a hard task not being influenced by the sound. Overall, all participants noticed the influence of the sounds interactively provided. This emerges not only from the several comments left, but also from results of the questionnaire item Q1. As illustrated in Fig. 5, the ratings of the extent with which the auditory feedback affected the way of walking are on average much higher than half of the Likert scale. Moreover, nine participants reported that different sounds adapted better to different emotions, stating that a particular emotional walking style was enhanced by a certain sound, or at the contrary, that a particular sound made hard to support a specific emotion.

However, the comments on the influence of each ground surface on each emotion were dissimilar. For instance, some subjects reported that the sound of snow enhanced the aggressive walking style, and others reported that it was harder to walk sadly or tenderly. Conversely, some of the participants noticed that snow helped in walking sadly because the sound brought to walk with a slower pace. This result could be due to cognitive factors, such as the recognition of the ground material, the degree of appreciation, the realism/artificiality, or the appropriateness of the sound felt by each walker. Indeed, the analysis of participants' comments revealed that the four sounds elicited different mental images. This is also reflected in the questionnaire results concerning the percentage of recognition of the simulated surfaces, where it is possible to notice that the four grounds were recognized with different extents. Specifically, the analysis of the answers revealed that some of the participants named the four grounds with different materials, including in several cases the impossibility to relate the sound to a ground material. However, as Fig. 5 shows, participants were good in classifying the sounds in the corresponding solid and aggregate surface type.

In summary, each participant adopted different strategies to cope with the provided sounds. On the one hand, some chose to ignore the sounds while others to second their influence. On the other hand, the mental images produced by each ground were not homogeneous across participants and this also affected the choices of how to walk when listening to the self-generated sounds.

The comparison between the performances of the two groups of participants revealed that the musical expertise did not yield a significant increase in the control of the body movements to render the emotions. Both groups produced emotions in a similar way. Similarly, no significant differences were found for the two groups either in the questionnaire results or in the reported comments.

Starting from all these results, a follow-up experiment was conducted in order to assess the ability of subjects to recognize the emotional intentions of a walker by listening to some of the recordings of the synthesized sounds produced during the first experiment. In particular, we were interested in verifying whether the emotions recognition was affected by the type of simulated ground material or by the type of musical expertise.

3 Recognition Experiment

3.1 Participants

Sixteen participants, who did not take part in the first experiment, were divided in two groups ($n = 8$). The first group was composed by musicians (3M, 5F), aged between 20 and 28 ($mean = 24.37$, $SD = 2.92$). They played different instruments (piano, trumpet, guitar, violin, viola, accordion) or were singers (soprano) with an average musical experience of 15.2 years. The second group was composed by non-musicians (4M, 4F), aged between 21 and 40 ($mean = 30.62$, $SD = 6.16$). Participants who belonged to this group never played any musical instrument nor took music lessons. However, they reported to listen to music on a daily basis. Participants had different nationalities (mostly Europeans). All participants reported neither hearing nor locomotion problems. They took on average about 38 minutes to complete the listening experiment, and were compensated for their participation.

3.2 Apparatus

The setup of the experiment was installed in an acoustically isolated laboratory and consisted of a laptop, a MIDI controller (Korg nanoKONTROL2), and the same sound card and headphones used in the production experiment. The laptop ran a graphical user interface that consisted of two parts. The fist part was composed by buttons, which were used to start and stop the trials. The second part was composed by five virtual sliders, each labelled with one of the five emotions, which were controlled by as many sliders of the MIDI controller. The instantaneous position of each slider was visualized on the screen, and could be varied between a minimum value indicated with "not at all" and a maximum value indicated with "very much".

Sounds were delivered through the headphones using the same sound card settings utilized during the first experiment. In this way participants to this second experiment could listen to exactly the same sounds experienced by the walkers in the first experiment.

3.3 Stimuli

Stimuli consisted of recordings of the synthesized footstep sounds produced by walkers in the production experiment. Twenty walking excerpts were selected from each group of participants, one for each combination of emotion and ground, for a total of 40 audio files. During the first experiment each subject produced two files for each combination of emotion and ground. To select one file among the 20 produced by the 10 participants of each group for each emotion-ground combination, such files were disposed in a bidimensional space, having as dimensions H2H and PK. Each file was defined by a pair (H2H, PK) according to the results of the analysis of the corresponding real footstep sounds performed during the first experiment. Subsequently, the file corresponding to the point in the bidimensional space with the smallest distance from the median value was chosen. For this purpose, the values of the two coordinates were first normalized between 0 and 1. With this selection criterion it was possible to choose the file closer to the median for each emotion and for each ground. Consequently the sound that in our hypothesis could be the most representative for each combination of emotion and ground was selected.

3.4 Procedure

Participants were asked to sit on a chair, to wear the headphones and to interact with the interface described in Section 3.2. The task consisted to rate each stimulus on the five scales expressing the emotions happiness, sadness, aggressiveness, tenderness, and emotionless, using the sliders of the MIDI controller. Each of the 40 stimuli was repeated twice for a total of 80 trials. When activated, each stimulus was looped with an interval of 2 seconds between the repetitions, so participants could listen to it as much as they wanted before giving an answer. Before passing to the subsequent stimulus, participants were asked to pull down all the sliders to the minimum value. When passed to the next stimulus they could not change the answer to the previous stimuli. Before the beginning of the experiment, participants familiarized with the setup and with the rating procedure practicing with five walking excerpts not included in the experiment.

At the end of the experiment, each of the participants took part to a survey consisting of three parts. The first part had to be completed before seeing the questions of the second part, and consisted of the following question: "*Describe which criteria you have used in order to evaluate the sounds*". In the second part participants were asked to rate on a seven point Likert scale [1 = not at all, 7 = very much] to what extent they had relied on the following criteria to perform the evaluations: Q1: volume of each footstep sound, Q2: duration of each footstep sound, Q3: frequency of each footstep sound, Q4: temporal distance between each footstep sound. The order of presentation of the questions was randomized. In the third part, participants were asked to recognize the type of sound (solid or aggregate) and to name the simulated surface materials. Finally, they were given the possibility to leave an open comment about their experience.

3.5 Results

The statistical analysis was inspired by the works presented in [9] and [12]. Firstly, for each of the selected recordings the mean values of H2H and PK of the sounds actually listened by the walkers were calculated according to the techniques described in Section 2.5.

Secondly, the relationships between both H2H and PK, and listeners' judgments were measured. For this purpose,

TABLE 2
Results of the Recognition Experiment: Kendall's Correlations between Each of the Two Acoustical Features and the Listeners Judgments for Each Emotion

	Metal		Wood		Gravel		Snow	
Emotion	**H2H**	**PK**	**H2H**	**PK**	**H2H**	**PK**	**H2H**	**PK**
Sadness	0.37***	-0.3***	0.25***	-0.2***	0.43***	-0.39***	0.35***	-0.37***
Tenderness	0.21***	-0.33***	0.16***	-0.2***	0.38***	-0.39***	0.18***	-0.26***
Neutralness	-0.1**	0.12**	0.04	-0.05	0.1	-0.1	0.17***	-0.24***
Happiness	-0.28***	0.21***	-0.13*	0.12**	-0.18***	0.17***	-0.13**	0.12**
Aggressiveness	-0.25***	0.33***	-0.41***	0.39***	-0.52***	0.54***	-0.42***	0.51***

** represents $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.*

the Kendall's correlations between each of the two acoustical features and the listeners' judgments were calculated. Thus, for example, the correlation between PK and sadness judgment indexes the extent to which the rating of sadness tends to increase or decrease when the sound level increases. Results are illustrated in Table 2. The Kendall's correlation test was adopted in place of the Pearson's correlation test used in [12], due to the recent evidence that subjective ratings should be transformed to ordinal representations for obtaining more reliable and generalisable models of affect [32].

Thirdly, in order to verify whether the emotions recognition was affected by the involved sounds, a Friedman Test was performed by considering the four materials for the absolute value of the correlations between the listeners' judgment and the two acoustic variables, calculated in the previous analysis.

Fourthly, to assess whether participants with musical expertise were better at recognizing the emotional walking intention we calculated, for each material and emotion, the phi correlation coefficient for binary classes considering for the two groups of participants the correctness in recognition (trials where the performed emotion was rated as the highest were considered correct). Results showed that all the correlations were low (all $\phi < 0.3$, with an average absolute value of 0.12 and SD = 0.08) showing, therefore, that musical expertise was not associated with a significant increase in the ability to recognize the emotions portrayed by the footstep sounds

As concerns the questionnaire, Fig. 6 shows the evaluations expressed by both the groups of participants for each questionnaire item. A Friedman Test was performed for each group of participants separately in order to assess if the differences between the ratings of the questionnaire items Q1, Q2, Q3, and Q4 were significant. A significant main effect was found for the group of the musicians, $\chi^2(3) = 19.47$, $p < 0.001$. The post-hoc analysis, performed by using the Wilcoxon-Nemenyi-McDonald-Thompson Test, revealed that to perform the evaluations participants relied less on the duration of the footstep sound criterion compared to the sound level of the footstep sound and the temporal distance between subsequent footstep (both $p < 0.05$). No significant main effect was found for the group of the non-musicians. In addition, no significant main effect was found between the two groups as result of a Mann-Whitney-Wilcoxon Test conducted for each questionnaire item.

3.6 Discussion

The results of the second experiment were in agreement with those reported in [8], [9] for a similar listening test

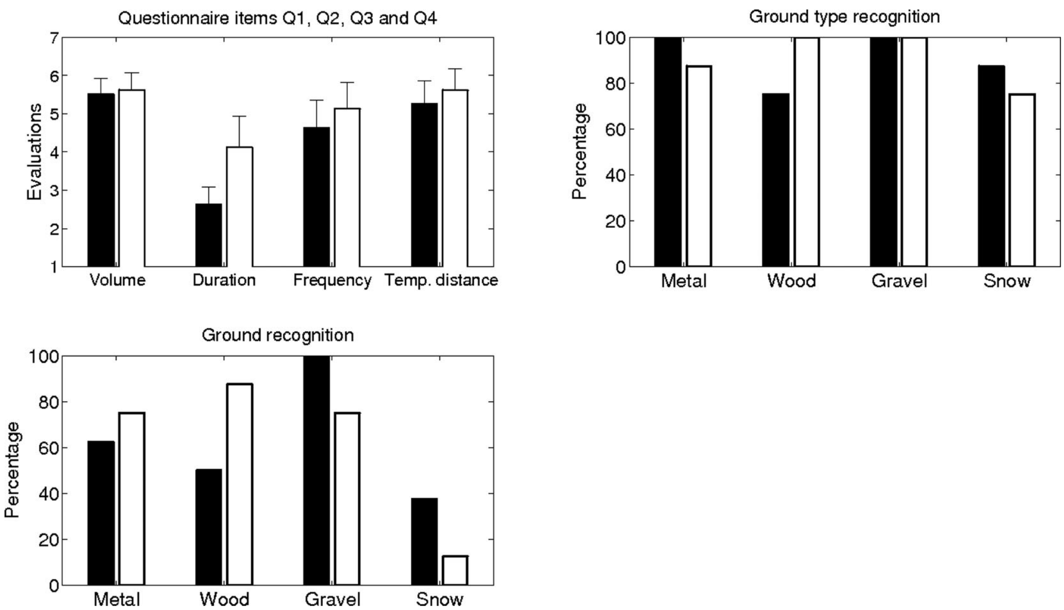

Fig. 6. Graphical representation of the mean and the standard error for participants' answers to questionnaire items Q1, Q2, Q3 and Q4 (top-left), percentage of recognition of the ground type (top-right), and percentage of recognition of the ground material (bottom-left). In black the ratings of the musicians and in white those of the non-musicians.

involving recordings of real footstep sounds, and were therefore consistent with previous results concerning the acoustical expression of emotions in music performance and vocality [11]. Specifically, sadness and tenderness were associated with low H2H and low PK, while happiness and aggressiveness were associated with high H2H and high PK.

The correlations suggest that the two acoustical features had considerable effects on the listeners' judgments of the emotional expression. With exception of the neutral emotion, all the correlations were statistically significant. Seventy five percent of these were medium or large according to the guidelines for interpretation of the effects provided by Cohen [33] (as used in [12]), i.e., small ($r \geq 0.1$), medium ($r \geq 0.3$), and large ($r \geq 0.5$), adapted to the Kendall's analysis as indicated in the [34], i.e., small ($\tau \geq 0.065$), medium ($\tau \geq 0.195$), and large ($\tau \geq 0.335$). As expected, and consistently with the results reported in [10], the neutral emotion was not associated with medium or large increases or decreases of the two acoustical features considered.

More importantly, the recognition of the walkers' emotional intentions was not influenced by the particular ground material sonically simulated. In addition, consistently with the results presented in [8], [9], the comparison between the recognition performances of the two groups of participants revealed that the musical expertise did not yield a significant increase in the ability to recognize emotions of a walker.

As far as the results of the questionnaire are concerned, listeners of both groups could classify the sounds in the corresponding solid and aggregate surface type, and recognition of the ground material was higher than chance for all sounds with exception of snow. Regarding the question about the criteria that participants relied on in order to evaluate the auditory stimuli, the majority of participants of both groups reported to have based their choices on walking speed and on the volume of the footstep sounds. Specifically, they reported to have associated walking excerpts characterised by slow paces and soft sounds to sad and tender emotions, and fast paces and loud sounds to happy and aggressive emotions. More than one participant commented about the difficulty to detect the happy emotion, and few of them reported to have based their choices also on the simulated ground material.

Regarding the questions in the second part of the questionnaire, the differences between the followed criteria were significant only for the group of musicians. This suggests that musicians were more able to distinguish the criteria which they relied on.

4 General Discussion

Our main aim in this research was to study whether and how synthesized footstep sounds interactively generated by walkers could affect their walking pace when systematically varying their intended emotion and the material of the ground.

The results of the first experiment revealed that different sonically simulated surfaces affected participants in different ways, as shown by Table 1. However, the influence of the sound manifested itself in more than one direction, not giving rise to statistically significant differences in the sound conditions, as illustrated in the overall results reported in Fig. 4. This result is analogous to the findings reported in [7], where no statistically significant differences between the sound conditions were found for a task similar to that of the present study.

An explanation for this result lies in the fact that participants interpreted the sounds in different ways, as emerged in the additional comments spontaneously delivered at the end of the experimental session. In general, participants reported to have the impression that some sounds better adapted to the different emotions, while others not. More importantly, they reported that the provided sounds had actually an influence on their performance, and as a consequence they decided either to control or to second them. A similar tendency was marginally emerged in [6]: two participants reported having ignored the interactive auditory feedback and indeed their performances turned out to be less influenced by the sounds compared to the rest of the group. Therefore, the findings of the first experiment show that walking with emotional intentions when provided with ecologically-valid interactive auditory feedback might collect very different results. The effect of the auditory feedback seems to depend on a participant basis: the individual propensity to be involved in the simulation seems to have an influence on the performance [35].

Therefore, these results did not confirm our original hypothesis that sound would have had an homogeneous influence among participants, as instead reported in [6]. However, the task involved in that study was very different from the one reported here. First of all, the walks were performed without the simulation of an emotional intention, i.e., participants were not forced to a particular gait style. Secondly, participants were requested to walk straight for 50 meters in each trial, so the covered distances were on the one hand greater, and on the other hand not interrupted by the back and forth procedure adopted in the present study. Thirdly, the experiment was conducted in an uncontrolled outdoor environment.

As for the two groups of participants, our hypothesis that musical expertise would have affected the walking performance was not confirmed.

As far as the second experiment is concerned, results showed that participants were capable of recognizing the emotional intentions of a walker when only acoustical information was available. Such an ability was not modulated by the ground material sonically simulated. In addition, the highest recognition performance was achieved for a non-emotional attribute of the sound, i.e., the typology of surface material. These results, therefore, lend support to the hypothesis reported in [8], [9] that there could be a separation on the perceptual use of acoustical parameters such that tempo and sound level are used by listeners to identify information related to the expression of emotions, while other parameters (e.g., spectral properties, sound duration, sound granularity) are used for the identification of the acoustical characteristics of the sound source. Moreover, analogously to the findings of [8], [9], listeners did not use music-related knowledge to recognize emotions in walking sounds.

The results of the two experiments replicated previous studies, both with regard to performers and listeners [7], [8], [9]: timing and sound level play a relevant role in emotional walking, both in production and in recognition. In both experiments, sadness and tenderness were associated with low H2H and low PK, happiness and aggressiveness were associated with high H2H and high PK, while the neutral emotion received average H2H and PK values smaller than sad and tender, and greater than happy and aggressive.

As reported in [8], [9], these two acoustical features are used similarly in both walking and music performance domains to express emotion. Moreover, from results of the two experiments it is possible to conclude that musicians and non-musicians reacted in similar ways to the provided stimuli. Indeed, no significant differences between the two groups of participants were found neither for production nor for the recognition of the emotions, showing that both production and recognition of emotions in walking comes before musical training. Therefore, taken together, our results lend support to the "motor origin hypothesis of emotional expression in music" [8], [9].

It has to be noticed that this study has limitations. Participants performed their walks in a laboratory environment and simulated their emotions. Nevertheless, the results in terms of timing and sound level are in line with those reported in studies on emotions in music and speech. In a real-life situation we could have observed more extreme values while in this laboratory controlled situation we could probably observe stereotypical behaviours, still possible to be classified by listeners.

5 Conclusion

This paper described two experiments investigating the role of sonically simulated ground materials in modulating auditory production and recognition of walks performed with emotional intentions. Two groups of participants, musicians and non-musicians, were involved in both the experiments. Participants' performances in both experiments were analyzed in terms of the temporal distance between adjacent steps and the sound level of each step. These two acoustical features were considered in place of other descriptors (e.g., spectral parameters) since they are those that can be strictly controlled by a walker [8], [9].

The results of the first experiment showed that the involved auditory feedbacks affected the pattern of emotional walking in different ways, although such an influence manifested itself in more than one direction. This behaviour was common to both the groups of subjects. The results of the second experiment showed the absence of an influence of the sound conditions on the recognition of the emotions from acoustic information alone. Also, the emotions were recognized independently from the level of musical expertise.

Our results suggest that tempo and sound level are two acoustical features important in both production and recognition of emotions in walking. This confirms results from a previous work on emotions in walking [9]. In addition, the similarities of the presented results with those reported in the music performance domain [11], as well as the absence of significant differences between the performances of musically trained and untrained participants foster the "motor origin hypothesis of emotional expression in music" formulated by Giordano et al. [8], [9].

In future works we plan to compare the proposed interactive sonifications with other techniques not valid from the ecological point of view [13], [14] currently used during motor therapy [3], [4], [5], [22].

Acknowledgments

This work was supported partly by a grant from the Danish Council for Independent Research awarded to Luca Turchet (Grant Nr. 12-131985), and partly by a grant from the Swedish Research Council awarded to Roberto Bresin (Grant Nr. 2010-4654). L. Turchet is the corresponding author.

References

[1] T. Hermann, A. Hunt, and J. Neuhoff, Eds., *The Sonification Handbook*. Logos Publishing House, Berlin, 2011.
[2] F. Fontana and Y. Visell, Eds., *Walking with the Senses. Perceptual Techniques for Walking in Simulated Environments*. Logos-Verlag, Berlin, 2012.
[3] Y. Baram and A. Miller, "Auditory feedback control for improvement of gait in patients with multiple sclerosis," *J. Neurological Sci.*, vol. 254, no. 1, pp. 90–94, 2007.
[4] Y. Baram and R. Lenger, "Gait improvement in patients with cerebral palsy by visual and auditory feedback," in *Proc. Virtual Rehab. Int. Conf.*, 2009, pp. 146–149.
[5] M. Schauer and K.-H. Mauritz, "Musical motor feedback (MMF) in walking hemiparetic stroke patients: Randomized trials of gait improvement," *Clinical Rehab.*, vol. 17, no. 7, pp. 713–722, 2003.
[6] L. Turchet, S. Serafin, and P. Cesari, "Walking pace affected by interactive sounds simulating stepping on different terrains," *ACM Trans. Appl. Perception*, vol. 10, no. 4, pp. 23:1–23:14, 2013.
[7] R. Bresin, A. DeWitt, S. Papetti, M. Civolani, and F. Fontana, "Expressive sonification of footstep sounds," in *Proc. Interactive Sonification Workshop ISon*, 2010, pp. 51–54.
[8] B. Giordano and R. Bresin, "Walking and playing: What's the origin of emotional expressiveness in music?" in *Proc. 9th Int. Conf. Music Perception Cognition*, 2006, p. 436.
[9] B. L. Giordano, H. Egermann, and R. Bresin, "The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity," *PLOS ONE*, vol. 9, no. 12, p. e115587, 2014.
[10] R. Bresin and A. Friberg, "Emotion rendering in music: Range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.
[11] P. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bull.*, vol. 129, no. 5, pp. 770–814, 2003.
[12] P. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception." *J. Exp. Psychology: Human Perception Perform.*, vol. 26, no. 6, p. 1797, 2000.
[13] W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological Psychology*, vol. 5, no. 1, pp. 1–29, 1993.
[14] W. Gaver, "How do we hear in the world? Explorations in ecological acoustics," *Ecological Psychology*, vol. 5, no. 4, pp. 285–313, 1993.
[15] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proc. IEEE Virtual Reality Conf.*, 2010, pp. 147–153.
[16] R. Bresin and S. Dahl, "Experiments on gestures: Walking, running, and hitting," in *The Sounding Object*, D. Rocchesso and F. Fontana, Eds., Edizioni mondo estremo, Florence, 2003, pp. 111–136.
[17] T. Eerola, V. Alluri, and R. Ferrer, "Emotional connotations of isolated instruments sounds," in *Proc. 10th Int. Conf. Music Perception Cognition*, 2008, pp. 483–489.
[18] R. Godøy, E. Haga, and A. Jensenius, "Playing "air instruments": Mimicry of sound-producing gestures by novices and experts," in *Gesture in Human-Computer Interaction and Simulation*. New York, NY, USA: Springer, 2006, pp. 256–267.
[19] C. Gaser and G. Schlaug, "Brain structures differ between musicians and non-musicians," *J. Neurosci.*, vol. 23, no. 27, pp. 9240–9245, 2003.

[20] K. Hyde, J. Lerch, A. Norton, M. Forgeard, E. Winner, A. Evans, and G. Schlaug, "Musical training shapes structural brain development," *J. Neurosci.*, vol. 29, no. 10, pp. 3019–3025, 2009.
[21] G. Lemaitre, O. Houix, N. Misdariis, and P. Susini, "Listener expertise and sound identification influence the categorization of environmental sounds." *J. Exp. Psychology: Applied*, vol. 16, no. 1, p. 16, 2010.
[22] M. Thaut and M. Abiru, "Rhythmic auditory stimulation in rehabilitation of movement disorders: A review of current research," *Music Perception*, vol. 27, no. 4, pp. 263–269, 2010.
[23] A. Lamontagne, J. Fung, B. J. McFadyen, and J. Faubert, "Modulation of walking speed by changing optic flow in persons with stroke," *J. Neuroeng. Rehab.*, vol. 4, no. 1, p. 22, 2007.
[24] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," in *Proc. Digital Audio Effects Conf.*, 2010, pp. 161–168.
[25] L. Turchet, "Custom made wireless systems for footstep sounds synthesis," *Appl. Acoust.*, vol. 83, pp. 22–31, 2014.
[26] F. Avanzini and D. Rocchesso, "Modeling collision sounds: Non-linear contact force," in *Proc. Digital Audio Effects Conf.*, 2001, pp. 61–66.
[27] P. Cook, "Physically informed sonic modeling (phism): Synthesis of percussive sounds," *Comput. Music J.*, vol. 21, no. 3, pp. 38–49, 1997.
[28] A. Gabrielsson and P. N. Juslin, "Emotional expression in music," in *Handbook of Affective Sciences*, R. J. Davidson, H. H. Goldsmith, and K. R. E. Scherer, Eds. London, U.K.: Oxford Univ. Press, 2003, pp. 503–534.
[29] P. N. Juslin and J. A. Sloboda, *Music and Emotion: Theory and Research*, P. N. Jusling and J. A. Sloboda, Eds., London, U.K.: Oxford Univ. Press, 2001, vol. 20, no. 3.
[30] L. Peltola, C. Erkut, P. Cook, and V. Valimaki, "Synthesis of hand clapping sounds," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1021–1029, Mar. 2007.
[31] A. Friberg, E. Schoonderwaldt, and P. N. Juslin, "Cuex an algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals," *Acta Acustica United With Acustica*, vol. 93, pp. 411–420, 2007.
[32] H. Martinez, G. Yannakakis, and J. Hallam, "Don't classify ratings of affect; rank them!" *IEEE Trans. Affective Comput.*, vol. 5, no. 3, pp. 314–326, Jul.-Sep. 2014.
[33] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York, NY, USA: Academic Press, 1988.
[34] A. Gilpin, "Table for conversion of Kendall's tau to Spearman's rho within the context of measures of magnitude of effect for meta-analysis," *Educ. Psychological Meas.*, vol. 53, no. 1, pp. 87–92, 1993.
[35] D. Scott, C. Stohler, C. Egnatuk, H. Wang, R. Koeppe, and J.-K. Zubieta, "Individual differences in reward responding explain placebo-induced expectations and effects," *Neuron*, vol. 55, no. 2, pp. 325–336, 2007.

Luca Turchet received the masters degrees in computer science in 2006 from the University of Verona, as well as in classical guitar and composition, in 2007 and 2009, respectively from the Music Conservatory of Verona. In 2013, he received the PhD degree in media technology from Aalborg University Copenhagen, where he currently works as a post-doc researcher funded by a three-years grant for independent research awarded by the Danish Research Council. His main research interests include sonic interaction design, multimodal systems, human-computer interaction, and perception.

Roberto Bresin received the MSc and DSc degrees in 1991 and 2000, respectively. He is a professor of media technology at the Department of Media Technology and Interaction Design (MID), School of Computer Science and Communication (CSC), KTH Royal Institute of technology, Stockholm, Sweden. His main research interests include emotion in sound and music performance, sound in interaction, sonification, and expressive music performance. He has published more than 140 papers in international conferences and peer-reviews journals. For more information please visit http://www.kth.se/profile/roberto/