Emotion Rendering in Auditory Simulations of Imagined Walking Styles

Luca Turchet and Antonio Rodà

Abstract—This paper investigated how different emotional states of a walker can be rendered and recognized by means of footstep sounds synthesis algorithms. In a first experiment, participants were asked to render, according to imagined walking scenarios, five emotions (aggressive, happy, neutral, sad, and tender) by manipulating the parameters of synthetic footstep sounds simulating various combinations of surface materials and shoes types. Results allowed to identify, for the involved emotions and sound conditions, the mean values and ranges of variation of two parameters, sound level and temporal distance between consecutive steps. Results were in accordance with those reported in previous studies on real walking, suggesting that expression of emotions in walking is independent from the real or imagined motor activity. In a second experiment participants were asked to identify the emotions portrayed by walking sounds synthesized by setting the synthesis engine parameters to the mean values found in the first experiment. Results showed that the involved algorithms were successful in conveying the emotional information at a level comparable with previous studies. Both experiments involved musicians and non-musicians. In both experiments, a similar general trend was found between the two groups.

Index Terms—Emotion rendering, walking, footstep sounds

1 Introduction

Research on auditory perception of foot-floor interactions has shown that footstep sounds can convey information not only about the walker's gender [1], posture [2], and identity [3], but also his/her emotional state [4].

Music shares with footstep sounds the fact to be a structured auditory stimulus characterized by properties such as rhythm and intensity. However, while the relation between emotions and music has been largely analyzed in the last decades (e.g., [5]), relatively recent are the studies on emotions and walking styles. Specifically, it has been shown that when walking with different emotional intentions humans make variations of timing and sound level in the same way as found in expressive music performance [6]. For example, it has been found that music performances communicating happiness and happy walking styles are characterized by a faster tempo/pace and louder sound level relative to a neutral style, while performances and walking patterns communicating sadness are characterized by slower tempo/pace and softer sound level. As far as timbre-related features are concerned, Tajadura-Jiménez et al. showed that increasing the amount of high frequency components of footstep sounds provided interactively while walking is effective in making subjects feel more aroused and positive [7]. This result is in accordance with results of previous studies on music performance relating high spectral centroids and/or high pitch to the happy emotion [8], [9].

Such analogies between music and footstep sounds, concerning the affective state of the performer/walker, may have a theoretical background in studies that analyze the relation between emotions, expression, and the human sensory-motor system. For instance, the embodied music cognition theory [10] considers the human body as the mediator between the internal musical representation and experience, and the external environment (including sounds); the "motor origin hypothesis of emotional expression in music" (MOH) [4] sustains that musicians and listeners make use of general movement knowledge when expressing and recognizing emotions in music.

Several research efforts have been dedicated to the algorithmic simulation of footstep sounds [11], [12], [13], [14], [15]. However, to date scarce is the research attention devoted to the simulation of footstep sounds expressing the emotional status of the walker. A preliminary effort in this direction is represented by the work presented in [16] where a model is proposed to synthesize interactively various attributes of the walker, floor, and foot-floor interaction. In this way the synthesis of the footsteps could be malleable to the emotional influence of the user controlling the model.

On the other hand, recent research has focused on how simulations of footstep sounds provided interactively with the footfalls of a walker can affect his/her walking kinematics in the context of simulated emotional walks. The study reported in [17] presented two experiments conducted to investigate the role of four sonically simulated surface materials (metal, wood, gravel, and snow) in modulating both production and recognition of walks performed with five emotional intentions (happy, sad, aggressive, tender, and neutral). The results of the first experiment showed that the involved auditory feedback affected the pattern of emotional

- *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.*
- *A. Rodà is with the Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131, Padova. E-mail: roda@dei.unipd.it.*

Manuscript received 29 Jan. 2015; revised 16 Oct. 2015; accepted 22 Oct. 2015. Date of publication 21 Jan. 2016; date of current version 6 June 2017. Recommended for acceptance by B.W. Schuller. 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.2016.2520924

walking in different ways, although such an influence was not homogeneous for all subjects. The results of the second experiment showed the absence of an influence of the simulated surface materials on the recognition of the emotions from acoustic information alone. Similar results were found in both experiments for musically-trained and -untrained participants. Those results suggested that in both production and recognition of emotionally expressive walking sounds tempo and sound level are important acoustical features. In addition, the similarities of the results with those reported in the music performance domain, as well as the absence of an influence of musical expertise lent support to the MOH.

In a different vein, research has also demonstrated that by exploiting a brain-computer interface humans are able to walk in virtual environments without any muscular activity but simply imagining the feet movements [18], suggesting that the imagination of the movements of the feet is a mental task closely related to that of real walking. Interestingly, the study reported in [19] showed that participants asked to either actually walk or imagine themselves walking to a previously seen target took almost exactly the same time in the two conditions. This and other results [20] provide support for the hypothesis that motor imagery shares the same neural mechanisms that are involved in motor control of actual actions [21]. Along the same line, research has recently shown that humans can reenact walking patterns according to those depicted in sequences of footstep sounds [22].

Importantly, various findings in neuroscience research have provided evidence that hearing the sound of an action not only can allow the understanding of associated actions by means of the activation of mirror neurons [23], but also activates brain networks responsible for the execution of the movement that caused that sound [23], [24], as well as evokes the motor plan that controls the limbs responsible for that sound production [25]. In addition, mirror neurons are though to be not only responsible to couple perception and actions, but also tot bind these with emotional expressions [26].

Starting from all the results reviewed above, a study was conducted on how emotions can be rendered and perceived when using synthetic footsteps sounds in a context not involving the physical locomotion. Our investigation aimed at providing guidelines for both the design and control of emotionally expressive computerized walking sounds that are more ecologically valid than sounds without performance variations. Such a research is especially relevant to designers of footstep sounds for videogames or virtual reality contexts where the auditory feedback associated with an avatar's foot-floor interactions can convey its emotional state.

This work is articulated in two experiments. The first experiment consisted in the production of emotionally expressive walking sounds according to imagined walking scenarios. It was conducted in a desktop configuration, (i.e., without the physical walk of the subjects), where participants could manipulate some parameters of the footstep sounds synthesizer described in [15], by moving some sliders on a purpose-designed Graphical User Interface. The main goal was to identify, for different emotions and for different combinations of surface materials and shoe types, the mean values and ranges of variation of two parameters of the footstep sounds (temporal distance between consecutive steps and sound level). The adopted methodology was inspired to the study reported in [8] where participants were asked to communicate different emotional expressions by manipulating the score of different musical pieces by means of seven musical variables. Those results allowed to identify, for each of the investigated emotions, the mean values and ranges of the involved musical variables. By asking participants to imagine to produce walking actions we also tested for similarities between the expression of emotions in real [4], [17] and imagined motor activity. Based on results of similarities between imagined and executed actions [20], [21], and in particular locomotion [19], we expected results to be in accordance with those reported in previous studies on real walking.

The second experiment consisted in a listening test where participants were asked to identify the emotions portrayed by walking sounds synthesized by setting the parameters of the synthesis engine to the mean values found in the first experiment. In particular, we were interested in verifying whether the emotion recognition was affected by the type of simulated surface material and shoe. Both experiments involved musically-trained and -untrained participants to test whether musical expertise could influence the quality of the results. An absence of differences between the performances of the two groups would be interpreted as lending support to the MOH.

2 Experiment 1

The main aim of the first experiment was to identify for different emotions and various combinations of surface materials and shoe types, mean and range of variation of two parameters of footstep sounds corresponding to the simulated walker's velocity and foot-floor impact force, i.e., respectively the temporal distance between two consecutive heel strikes (heel-to-heel, H2H) and the sound pressure level measured at its peak (PK). These parameters were chosen because in previous research were proven to be the most salient acoustical features involved when producing walks with emotional intentions [4], [17]. Such a research also aimed at verifying whether the experimental results were consistent with previous findings that revealed similarities between the expression of emotions with a non-musical everyday motor activity such as walking and the musical expression of emotions as reported in [4] and [17]. This not only would lend support to the MOH, but also would suggest that the expression of emotions in walking is independent from the level of abstraction of the motor activity (real or imagined).

In addition, since participants were asked to imagine themselves walking at a self selected speed we searched for eventual correlations between their collected anthropomorphic features (height, weight, and foot length) and the values of the two investigated parameters in each emotion. Such correlations were not expected to be present for H2H, in accordance with findings reported in [27] that showed that H2H is not correlated with height and weight when humans walk at a self selected speed without any emotional intention. As a consequence, the validity of this hypothesis for the results of the neutral emotion condition could be interpreted as a further prove of the similarity of the mechanisms underlying real and imagined motor activity. Moreover, if its validity extends also to the other emotions as well as to the PK parameter, this

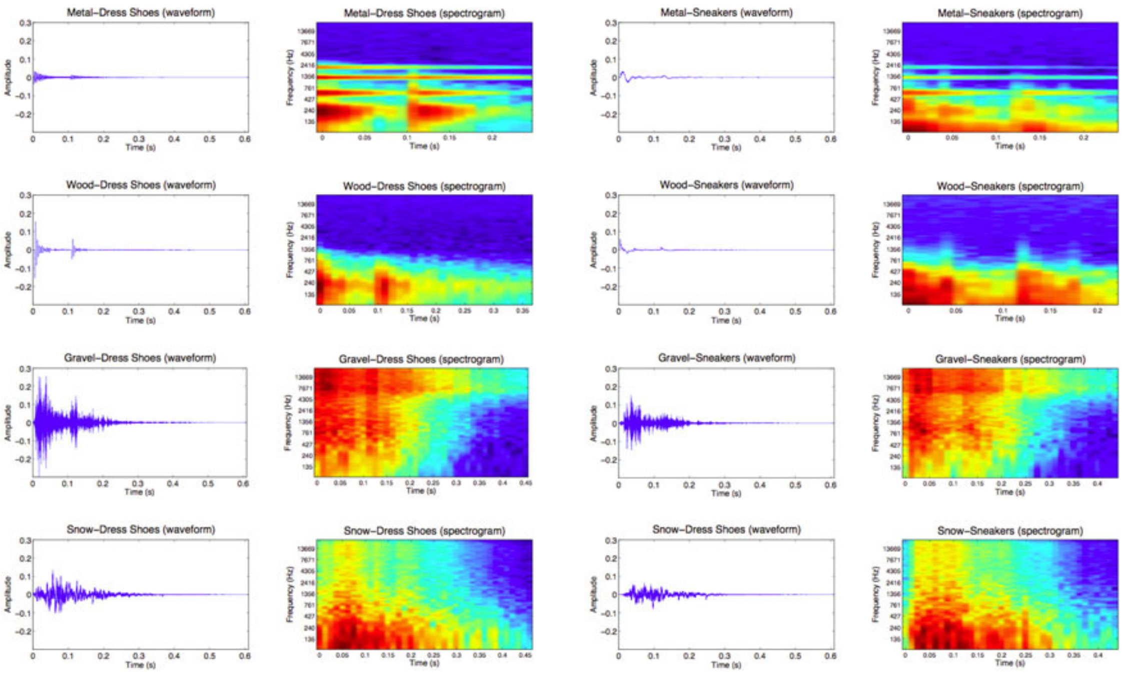

Fig. 1. Typical waveforms and spectrogram (with log frequency axis) of the eight combinations of shoe type and surface materials.

would allow to conclude that the production of emotionally expressive walking sounds in the context of passive sensory motor activity is independent from the anthropomorphic features of the person imagining the simulated walks.

2.1 Participants

Twenty participants were divided in 2 groups (n = 10) to perform the experiment. The first group was composed by musicians (9M, 1F), aged between 22 and 44 (mean = 32.3, SD = 8.23), with an average musical experience of 20.7 years. The second group was composed by non-musicians (5M, 5F), aged between 26 and 32 (mean = 29.1, SD = 1.91). Participants who belonged to this group never played any musical instrument nor took music lessons. All participants reported neither hearing nor locomotion problems.

2.2 Apparatus

The setup of the experiment was installed in a silent room and consisted of a laptop (Macbook Pro) and a closed headphone set (Sennheiser PXC 450). The laptop run the footstep sound synthesizer described in [15] and a graphical user interface, both realized in Max/MSP sound synthesis and multimedia real-time platform. The former allows to synthesize different combinations of surface materials, shoe types, foot-floor interactions and anthropomorphic features of the walker. The synthesis is based on physical, physically informed, and psychologically informed models. The ecological validity of the simulated sounds was assessed in [28] and [15]. The latter consisted of buttons to start and stop the trials, a label displaying the emotion to be rendered in each trial, and two virtual sliders labelled "velocity" and "volume" allowing to vary H2H and PK respectively.

2.3 Stimuli

The synthesis engine was set to simulate footsteps on four surface materials (two solids, wood and metal, and two aggregates, snow and gravel) performed by a genderless walker wearing two types of shoes (dress shoes and sneakers). These materials and shoe types were chosen because the resulting eight combinations gave rise to a rather comprehensive palette of footstep sounds, which had a variety of both temporal and spectral features. The surface materials were the same utilized in the experiment reported in [17]. They are characterized by different features in terms of duration, amplitude, onset type, temporal evolution, and spectrum. They were proven to be easy to recognize, as well as to classify in the corresponding solid and aggregate surface typology [28] and to express different degrees of compliance. Similarly, the two types of shoes were characterized by two levels of sole hardness (hard and medium) that were simulated as giving rise to sounds having different properties in terms of onset, peak, and duration when impacted on a same material. Specifically, sneakers were simulated with longer onset, lower peak and shorter duration compared to dress shoes [15]. Fig. 1 illustrates the waveforms and the spectra of the eight combinations of shoe type and surface material.

In the experiment, participants could continuously vary the H2H parameter in the range of [319, 1,276] ms. A logarithmic scale for the slider associated to the H2H parameter was utilized according to the logarithmic perception of time changes by humans (Weber-Fechner law). The time range was selected to be one time slower and one time faster than a mean value of 638 ms, which could fit for a genderless walker. Such a value was calculated by averaging the values of H2H collected from 20 participants (10 M and 10 F) of the interactive experiment reported in [17], who performed walks with the neutral emotion intention while listening the sounds of the same four surface materials here involved. The time range was checked in an informal listening session in which the authors manipulated the time of the experiment stimuli. H2H period values falling outside of the proposed range resulted unrealistic for a walk: too great values of H2H may break the perception of a continuous walk, whereas too low values may give the perception of a run.

In the experiment, participants could continuously vary the sound level by moving the corresponding slider in the range of [−24, 24] dB to deviate from the nominal sound level of each stimulus. For the four stimuli involving the dress shoes, the nominal sound level was chosen by averaging the peak value of the sound level selected by the 20 participants of the interactive experiment reported in [17], while performing the walks with the neutral emotional intention on the four surface materials wearing dress shoes. Specifically, the nominal sound level of the peak value for the stimuli was 62.5, 65.4, 66.1, and 64.9 dB for metal, wood, gravel, and snow respectively. For the four stimuli involving the sneakers, the nominal sound level was 58.4, 57.5, 60.1, and 61 dB for metal, wood, gravel, and snow respectively. The PK range was checked in an informal listening session in which the authors manipulated the PK of the experiment stimuli. PK values falling outside this range would result unrealistic for a walk: too high values of PK would break the perception of a step while too low values would not be clearly audible.

2.4 Procedure

Participants were presented with written instructions. They were asked to sit on a chair, to wear the headphones, and to interact with the graphical interface using a mouse as a control device. The task consisted in manipulating the "velocity" and "volume" sliders, to produce five emotional intentions (happy, aggressive, tender, sad, and neutral) for each of the eight types of stimuli, for a total of 40 trials. For each emotion, participants were instructed to adjust the two sliders as they were walking in a specific situation 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"*

These situations were the same involved in [17]. 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 [29]), and because they cover the four quadrants of the two-dimensional Activity-Valence space [5]. 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 [4]. 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 [8].

Trials were presented in randomized order. Each trial was presented once, but participants were allowed to experience each stimulus as much time as they needed to make their choice. Before the beginning of the experimental session participants practiced 10 trials to familiarize with the setup. Those trials were not included in the experiment and consisted of four stimuli resulting from the combination of the two types of shoes and two surface materials, concrete and forest underbrush. The 10 trials were randomly chosen among the 20 trials resulting from the combination of the four stimuli and five emotions. Participants were not informed about the involved types of shoes and surface materials.

At the end of the experimental session participants were asked to fill a questionnaire gathering information regarding some anthropomorphic features (height, weight, foot length) and the level of musical expertise. In addition, participants were asked to name the simulated surface materials and shoe types. Finally, they were given the possibility to leave an open comment about their experience.

2.5 Results

Fig. 2 illustrates the experimental results for the two investigated parameters and for the two groups of participants. Tables 1 and 2 report the means, their standard error, and confidence interval for the two investigated parameters averaged between the two groups of participants.

Statistical analysis was performed on the collected data by means of repeated measures ANOVAs. Before running all the ANOVAs, we checked for the normality of the data distribution by means of a Shapiro-Wilk test, and a Mauchlys test was applied for verifying if the assumption of sphericity had been met for the investigated factors. All post-hoc analyses were performed by using Tukey's procedure (p-value was set at a significant $p < .05$). Both the collected H2H and PK were subjected to a four-way analysis of variance having five levels of emotion (happy, aggressive, tender, sad, and neutral), four levels of material (metal, wood, gravel, and snow), two levels of shoe type (dress shoes and sneakers), and two levels of musical expertise (musicians and non-musicians).

A significant main effect of H2H was found for musical expertise, $F(1,18) = 4.467$, $p < .05$, as well as for emotion, $F(4,72) = 130.207$, $p < .001$. The pairwise comparison showed that H2H was significantly shorter for aggressive compared to neutral, tender, and sad ($p < .001$), for happy compared to neutral, tender, and sad ($p < .001$), and neutral compared to sad ($p < .05$). There was not a significant main effect either for factors material or shoe type. The sole significant interaction effect was between emotion and material, $F(12,216) = 2.812$, $p < .01$ (see Fig. 3).

A significant main effect of PK was found for musical expertise, $F(1,18) = 5.123$, $p < .05$, as well as for emotion, $F(4,72) = 57.096$, $p < .001$. The pairwise comparison showed that PK was significantly greater for aggressive compared to neutral, tender, and sad ($p < .001$), for happy compared to neutral ($p < .05$), tender, and sad (both $p < .001$), and neutral compared to sad ($p < .05$). A significant main effect of PK was found for material, $F(3,54) = 63.838$, $p < .001$. The pairwise comparison showed that PK was significantly greater for wood compared to snow ($p < .05$). A significant main effect of PK was found for shoe type, $F(1,18) = 6.732$, $p < .05$. The sole significant interaction effect was between shoe and material, $F(3,54) = 25.4$, $p < .001$ (see Fig. 4).

The Pearson's correlation between the two investigated parameters was computed, showing a significantly negative

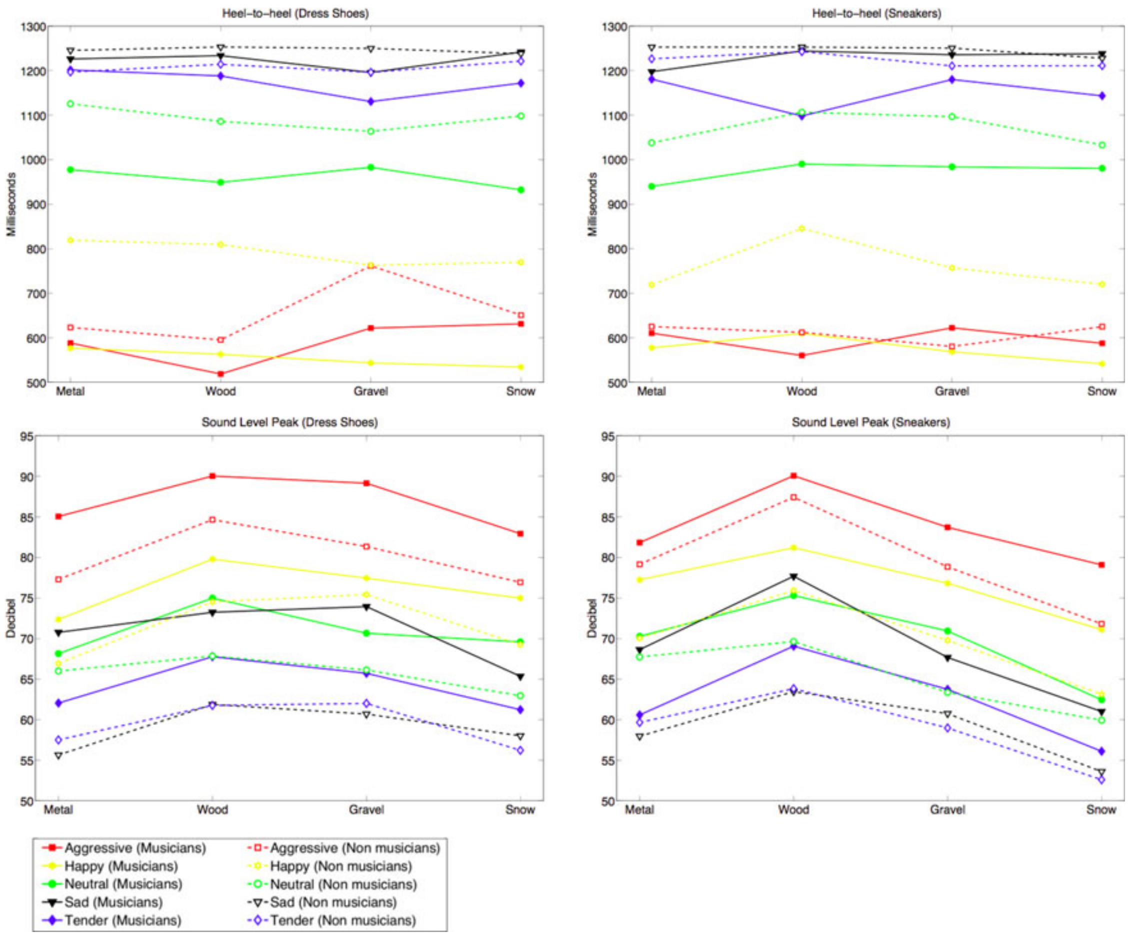

Fig. 2. Graphical representation of the results of the experiment 1 for each emotion, surface material, shoe type, and participants group.

correlation between H2H and PK ($r = -0.56$, $p = \leq .001$; $df = 798$). A linear mixed-effects model analysis was performed for each emotion separately considering the possible correlations between the two investigated parameters and the collected participants' anthropomorphic features (weight, height, foot length). Such analysis revealed that

TABLE 1
Means, Standard Errors and Confidence Intervals for the Two Variables H2H and PK as Rendered for the Five Emotions by Participants and for the Trials Involving Dress Shoes

Material	Emotion	H2H (ms)			PK (dB)		
		Mean	Std. Error	95% Confidence interval	Mean	Std. Error	95% Confidence interval
Metal	Aggressive	605	44	[512, 698]	81.1	2.3	[76.2, 86.1]
	Happy	697	64	[562, 833]	69.6	2.2	[64.9, 74.3]
	Neutral	1,051	31	[985, 1,117]	67	1.5	[63.7, 70.4]
	Tender	1,199	21	[1,154, 1,243]	59.7	1.8	[55.8, 63.6]
	Sad	1,235	11	[1,210, 1,260]	63.2	2.7	[57.4, 68.9]
Wood	Aggressive	557	33	[487, 626]	87.3	2.3	[82.3, 92.3]
	Happy	686	54	[571, 800]	77.1	2.1	[72.6, 81.6]
	Neutral	1,017	39	[935, 1,099]	71.4	1.5	[68.1, 74.6]
	Tender	1,201	16	[1,165, 1,236]	64.7	1.7	[61.1, 68.3]
	Sad	1,243	7	[1,228, 1,258]	67.5	2.6	[61.9, 73.1]
Gravel	Aggressive	691	47	[591, 791]	85.2	2.1	[80.7, 89.7]
	Happy	653	58	[530, 776]	76.4	2.3	[71.6, 81.2]
	Neutral	1,023	38	[942, 1,104]	68.3	1.4	[65.3, 71.3]
	Tender	1,163	25	[1,110, 1,216]	63.8	1.8	[59.9, 67.7]
	Sad	1,222	18	[1,184, 1,261]	67.3	2.3	[62.3, 72.2]
Snow	Aggressive	641	52	[530, 751]	79.9	2.2	[75.2, 84.6]
	Happy	651	57	[531, 772]	72.1	2.3	[67.1, 77]
	Neutral	1,015	38	[935, 1,094]	66.2	2.2	[61.6, 70.9]
	Tender	1,196	17	[1,160, 1,233]	58.7	1.6	[55.1, 62.2]
	Sad	1,240	7	[1,223, 1,256]	61.6	2.1	[57.1, 66.2]

TABLE 2
Means, Standard Errors and Confidence Intervals for the Two Variables H2H and PK as Rendered for the Five Emotions by Participants and for the Trials Involving Sneakers

Material	Emotion	H2H (ms)			PK (dB)		
		Mean	Std. Error	95% Confidence interval	Mean	Std. Error	95% Confidence interval
Metal	Aggressive	617	44	[524, 711]	80.4	1.5	[77.3, 83.6]
	Happy	648	48	[546, 750]	73.6	2.1	[69.1, 78.1]
	Neutral	989	38	[909, 1,068]	68.9	1.4	[65.9, 71.9]
	Tender	1,203	16	[1,169, 1,237]	60.1	2	[55.8, 64.4]
	Sad	1,225	18	[1,186, 1,264]	63.3	2.6	[57.7, 68.8]
Wood	Aggressive	586	44	[493, 678]	88.7	1.5	[85.5, 91.9]
	Happy	727	56	[608, 846]	78.5	2.1	[74, 83]
	Neutral	1,048	31	[981, 1,115]	72.4	1.5	[69.1, 75.7]
	Tender	1,170	27	[1,113, 1,227]	66.4	1.9	[62.3, 70.5]
	Sad	1,248	6	[1,234, 1,262]	70.5	3.3	[63.6, 77.4]
Gravel	Aggressive	601	37	[523, 679]	81.2	2.1	[76.8, 85.7]
	Happy	662	50	[557, 768]	73.2	2.4	[68.1, 78.4]
	Neutral	1,040	28	[980, 1,099]	67.1	1.3	[64.2, 70]
	Tender	1,195	16	[1,160, 1,229]	61.3	1.8	[57.5, 65.1]
	Sad	1,243	7	[1,227, 1,258]	64.2	2.3	[59.2, 69.1]
Snow	Aggressive	606	47	[506, 705]	75.4	2.2	[70.8, 80]
	Happy	630	46	[533, 728]	67.1	2.5	[61.8, 72.4]
	Neutral	1,006	45	[911, 1,102]	61.1	1.3	[58.3, 64]
	Tender	1,177	23	[1,128, 1,226]	54.3	1.8	[50.4, 58.2]
	Sad	1,232	12	[1,206, 1,258]	57.3	2.4	[52.2, 62.3]

participants' anthropomorphic features were not linearly related either to H2H or PK.

As far as the identification of ground materials is concerned, considering both correct and incorrect answers results of both experiments revealed that participants perceived clearly the difference between solid and aggregate surfaces, and this result is in accordance with the findings reported in [15] and in a previous study using the same footstep sounds engine [28]. Similarly, results on the identification of the shoe types in both experiments paralleled those reported in [15].

2.6 Discussion

This experiment allowed to identify for different emotions and for different combinations of surface materials and shoe types, the mean values and ranges of variation of PK and H2H. These results can be practically utilized to build a synthesizer for emotionally expressive footstep sounds.

What emerges from results is that the two investigated parameters were manipulated in the same direction as reported in previous research on real walks performed with emotional intentions [4], [17], which in turn were consistent with studies on emotional expressive music performance [8], [29]. In agreement with [4] and [17], H2H showed a negative correlation with PK. This result reflects reports from previous studies on emotionally expressive music performance [6], [8]: when a piece of music is played with faster tempo, it is also played louder. However, studies on musical performance that involved a wider set of adjectives showed that this result could depend by the use of only five basic emotions. For example, piano performances suggested by adjectives such as heavy, solemn, or obscure are characterized by medium-high intensity and low tempo (corresponding to high H2H) [30].

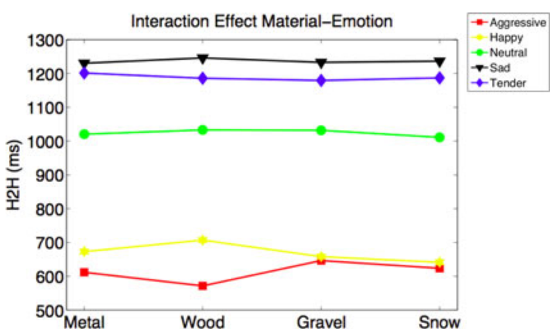

Fig. 3. Interaction effect between emotion and material for H2H.

Musical expertise yielded to significantly different results, contrarily to the findings reported in [4] and [17]. In Fig. 2 it is possible to notice how for each emotion musicians' choices of H2H and PK were respectively lower and greater compared to those of non-musicians. Authors are not able to find a plausible explanation for this different behaviour. The greatest difference between the two groups of participants concerns the average values of PK for the sad emotion for all the combinations of surface material and shoe type. Moreover, non-musicians chose very similar values of H2H and PK both for sad and tender walks, showing that they were not able to differentiate these two emotions, diversely by musicians which chose higher PK values for sad in respect to tender. This result could be explained by a greater ability of musicians in controlling music-related parameters such as tempo and intensity. Therefore, these results lend support to the MOH only in part.

Participants' choices of H2H were independent from the type of shoe sonically simulated for each emotion. Conversely, they depended on the surface material as far as aggressiveness and happiness are concerned: solids were assigned with faster speeds than aggregates for the aggressive emotion, while slower for the happy one (see Fig. 3). Moreover, metal and wood received more dissimilar speeds

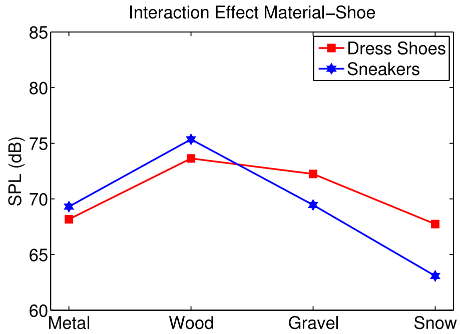

Fig. 4. Interaction effect between shoe and material for PK.

compared to snow and gravel. This result supports the idea that subjects were less able to differentiate aggressive from happy emotions on aggregate materials than on solid surfaces.

Participants' choices of PK depended both on shoe type and surface material. As far as the material is concerned, wood received values of PK significantly greater than snow. It is interesting to notice that these two materials present, for both the involved simulations of shoe type, respectively the longest and shortest durations among the four materials. It is well known in the literature that for sounds of duration lower than 1 second (as it is the case of footstep sounds), the lower the duration the lower the perceived loudness for the same value of PK [31]. Therefore, it might be that to compensate this bias in perception participants adjusted the PK of wood to greater values compared to snow. On the other hand, the interaction effect depicted in Fig. 4 shows that dress shoes were assigned with volumes greater than sneakers in presence of aggregates materials, and lower volumes in presence of solid materials. This result is rather counter-intuitive: one would expect to have for solids greater PK ratings for dress shoes compared to sneakers since hard sole shoes produce louder sounds compared to soft sole shoes when impacted with the same force on the floor. An explanation might be found in the impulsive nature of the sound resulting from shoes-impacts on solid materials compared to the sound's granular structure of aggregates. As a matter of fact, previous studies have shown that these two surface materials typologies are processed in very different ways [28], [32], [33].

The hypothesis of similarities between real and imagined motor activity in expression of emotions was confirmed. First, as expected no correlations were found between the values of the two investigated parameters and participants' height, weight, and foot length. This result is in accordance with findings reported in [27] for H2H in the neutral emotion condition, and extends also to all the other four simulated emotions and PK. Second, H2H and PK ratings were consistent for each emotion with those found in previous studies involving real walking (e.g., sad walks received lower ratings of H2H and PK compared to happy walks) [4], [17]. However, it has to be noticed that these studies are comparable with the present one only at level of general trend but not in terms of the mean values of the two investigated parameters. Indeed, on the one hand the study reported in [4] involved seven participants walking on a linoleum floor, wearing different types of shoes, and not instructed to walk with the same scenarios described in Section 2.4; on the other hand the study reported in [17] involved interactive simulations of footstep sounds that differed from the sounds utilized here and above all had an influence on the participants' walking patterns. Further research is needed to assess to what extent real and imagined emotionally expressive walking styles are similar.

3 Experiment 2

The aim of the second experiment was to verify to what extent the emotional intentions portrayed by the sounds generated by setting the synthesis engine with the mean values resulting from the first experiment could be recognized in a listening test, and whether the recognition was moderated by the type of surface material and type of shoe sonically simulated. In fact, the mean values obtained with the first experiment with the various combinations of surface materials and shoe types represent a sort of expressive palette for rendering different emotional intentions by means of synthesized footstep sounds. The results of this experiment represent therefore a validation of such a palette. As in the first experiment, we used two groups of participants, musicians and non-musicians. The rationale to use these two groups was to assess whether the listeners' expertise yelled different results. Based on the results of similar listening tests reported in [4] and [17], we expected that listeners would not have used music-related knowledge to recognize emotions in walking sounds.

3.1 Participants

Twenty participants, all of whom were not involved in the first experiment, were divided in 2 groups (n = 10) to perform the experiment. The first group was composed by musicians (6M, 4F), aged between 20 and 42 (mean = 25.5, SD = 6.13), with an average musical experience of 15.2 years. The second group was composed by non-musicians (5M, 5F), aged between 20 and 28 (mean = 23.6, SD = 3.02). Participants who belonged to this group never played any musical instrument nor took music lessons. All participants reported neither hearing nor locomotion problems.

3.2 Apparatus

The setup of the experiment was installed in a silent room and involved the same apparatus and footstep sounds synthesizer utilized in the first experiment, with exception of the graphical user interface. This consisted of two parts. The first 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 (happy, sad, aggressive, tender, and neutral). The sliders corresponded to five values scales and were controlled by the mouse device. The position of each slider could be continuously varied between a minimum value indicated with "not at all" and a maximum value indicated with "very much".

3.3 Stimuli

Stimuli consisted of 40 audio excerpts generated by setting the sound synthesis engine to simulate the same 40 combinations of emotion, surface material, and shoe type involved in the first experiment. Each emotion was

TABLE 3
Subjects' Rating on the Five Evaluation Scales of the Stimuli Generated for the Experiment 2 [Median (IQR)]

Emotion portrayed by the stimuli	Evaluation scale				
	Aggressive	Happy	Neutral	Sad	Tender
Aggressive	**6.16** (2.23-8.34)	0 (0-5.35)	0 (0-2.09)	0 (0-0)	0 (0-0)
Happy	0 (0-4.21)	1.14 (0-5.43)	**3.32** (0-7.59)	0 (0-0)	0 (0-0)
Neutral	0 (0-0)	0 (0-0.41)	**3.72** (0-6.97)	.69 (0-4.99)	1.22 (0-4.44)
Sad	0 (0-0)	0 (0-0)	1.5 (0-6.13)	**3.06** (0-6.28)	2.03 (0-5.08)
Tender	0 (0-0)	0 (0-0)	1.71 (0-5.75)	2.29 (0-6.29)	**2.65** (0-6.04)

The values are grouped for the emotion portrayed by the stimuli according to the results of experiment 1.

rendered by using the mean values of PK and H2H reported in Tables 1 and 2. Each audio excerpt consisted of six steps.

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 in rating each stimulus on the five scales expressing the emotions happiness, sadness, aggressiveness, tenderness, and emotionless. Each of the 40 stimuli was repeated twice for a total of 80 trials, which were presented in randomized order. When activated, each stimulus was looped with an interval of 2 seconds between the repetitions, so as participants could listen to it as much as they wanted before giving an answer. When the answer was chosen, the sound stopped and all the sliders of the interface were automatically set to the minimum value. When passed to the next stimulus participants could not change the answer to the previous stimuli. Before the beginning of the experiment, they familiarized with the setup and with the rating procedure practicing with five walking excerpts not included in the experiment. These consisted of six footstep sounds produced on concrete and on forest underbrush by using dress shoes. The values of H2H and PK of these stimuli were those utilized for the wood and gravel respectively reported in Table 1.

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 second part, and consisted of the following assignment: "*Describe which criteria you used to evaluate the sounds*". In the second part, participants were asked to rate on a visual analogue scale (VAS) 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: pitch 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 name both the simulated surface materials and the type of shoes. Finally, they were given the possibility to leave an open comment about their experience.

3.5 Results

Every subject evaluated each of the 80 stimuli along 5 continuous evaluation scales. In order to obtain more reliable and generalisable results, the ratings were treated as ordinal (ranked) values [34] and non parametric statistical tests were applied.

First of all, we assessed whether the subjects were able to recognize the emotion portrayed by the stimuli, confirming the results of experiment 1. In other words, we wanted to verify the hypothesis that the walking sounds generated following the values of PK and H2H reported in Tables 1 and 2 were effective in conveying the emotions listed in the same tables. Table 3 shows the median and the interquartile range (IQR) of the subjects' ratings on the five evaluation scales, grouping the values on the basis of the emotion portrayed by the stimuli. According to a Kruskal-Wallis one-way analysis of variance by ranks, a significant effect of the evaluation scales was found for every emotion portrayed by the stimuli ($\chi^2(4) = 247.24$, $p < .001$). Table 4 shows the results of a non parametric multiple comparison test, carried out following the Tukey's method (p-value was set at a significant $p < .05$) [35]. According to the median values, the subjects rated the aggressive stimuli as more aggressive (median = 6.16) and the pairwise comparison showed that this value was significantly greater ($p < .001$) than happy,

TABLE 4
Pairwise Comparisons between the Median Ratings of Experiment 2 (*p*-Value)

Comparison	Emotion portrayed by the stimuli				
	Aggressive	Happy	Neutral	Sad	Tender
Aggressive-Happy	$<.001$	.105	.208	1.000	.067
Aggressive-Neutral	$<.001$	$<.001$	$<.001$	$<.001$	$<.001$
Aggressive-Sad	$<.001$	$<.001$	$<.001$	$<.001$	$<.001$
Aggressive-Tender	$<.001$	$<.001$	$<.001$	$<.001$	$<.001$
Happy-Neutral	$<.001$	$<.001$	$<.001$	$<.001$	$<.001$
Happy-Sad	$<.001$	$<.001$	$<.001$	$<.001$	$<.001$
Happy-Tender	$<.001$	$<.001$	$<.001$	$<.001$	$<.001$
Neutral-Sad	$<.05$	$<.001$	$<.001$	.092	.554
Neutral-Tender	$<.001$	$<.001$	$<.001$	.996	.260
Sad-Tender	$<.001$	.906	1.000	.137	.994

TABLE 5
Subjects' Ratings of the Experiment 2

Material	Emotion portrayed by the stimuli	Evaluation scale				
		Aggressive	Happy	Neutral	Sad	Tender
Gravel	Aggressive	**6.79** (1.8-8.1)	.24 (0-5.48)	0 (0-0.43)	0 (0-1.37)	0 (0-0)
	Happy	0 (0-2.55)	**5.25** (1.2-6.82)	3.4 (0-7.59)	0 (0-0)	0 (0-0)
	Neutral	0 (0-0)	0 (0-0.74)	**4.61** (0-7.14)	0 (0-3.32)	2.77 (0-5.4)
	Sad	0 (0-0)	0 (0-0)	**3.52** (0.51-6.69)	2.33 (0-4.39)	.81 (0-3.87)
	Tender	0 (0-0)	0 (0-1.74)	2.99 (0-5.97)	.83 (0-4.57)	**3.36** (0-6.59)
Metal	Aggressive	2.69 (0.43-6.6)	**4.17** (0-7.41)	0 (0-1.87)	0 (0-0)	0 (0-0)
	Happy	0 (0-2.42)	**5.03** (1.2-7.07)	1.69 (0-5.9)	0 (0-0)	0 (0-0)
	Neutral	0 (0-0)	0 (0-4.51)	**6.08** (2.03-7.99)	0 (0-0.08)	0 (0-4)
	Sad	0 (0-0)	0 (0-0)	3.11 (0-6.78)	.78 (0-2.95)	**3.16** (0-6.39)
	Tender	0 (0-0)	0 (0-1.34)	**4.74** (0-6.97)	0 (0-3.35)	2.57 (0-6.15)
Snow	Aggressive	**9.22** (8.38-10)	0 (0-1.46)	0 (0-0)	0 (0-0.15)	0 (0-0)
	Happy	**6.84** (4.27-8.55)	0 (0-4.71)	0 (0-0.46)	0 (0-1.92)	0 (0-0)
	Neutral	0 (0-2.88)	0 (0-0)	0 (0-5.28)	**3.59** (0-7.47)	0 (0-2.1)
	Sad	0 (0-0.64)	0 (0-0)	0 (0-3.66)	**5** (1.35-6.87)	0 (0-2.48)
	Tender	0 (0-0)	0 (0-0)	.01 (0-5.09)	**5.25** (3.18-7.32)	1 (0-4.19)
Wood	Aggressive	**6.06** (1.66-7.46)	1.94 (0-5.41)	0 (0-3.97)	0 (0-0)	0 (0-0)
	Happy	0 (0-0.08)	0 (0-0.18)	**7.23** (4.39-8.27)	0 (0-0.66)	0 (0-1.64)
	Neutral	0 (0-0)	0 (0-0)	**4.71** (0-6.47)	0 (0-4.72)	2.97 (0-5.57)
	Sad	0 (0-0)	0 (0-0)	3.21 (0-6.93)	**3.57** (0-6.86)	3.29 (0.76-6.84)
	Tender	0 (0-0)	0 (0-0)	2.42 (0-5.91)	1.88 (0-5.23)	**3.31** (0-7.69)

The values [Median (IQR)], grouped for surface material, refer only to trials involving dress shoes.

neutral, tender, and sad (median = 0); happy stimuli were rated as neutral (median = 3.32), a value significantly greater ($p < .001$) than happy (median = 1.14), aggressive, tender, and sad; neutral stimuli were rated as neutral (median = 3.72), a value significantly greater ($p < .001$) than the other evaluation scales; sad stimuli were rated as sad (median = 3.06) but the value was not significantly different from tender (median = 2.03, $p = .137$) and neutral (median = 1.5, $p = .092$); finally, tender stimuli were rated as tender (median = 2.65) but this value was not significantly different from sad (median = 2.29, $p = .994$) and neutral (median = 1.71, $p = .554$).

In order to emphasize the effect of the factors surface material and shoe type on the recognition task, the subjects' rates were analyzed separately for dress shoes (see Table 5) and sneakers (see Table 6), grouping the values by surface material. The aggressive stimuli were generally well recognized in every condition but metal-dress shoes: the median

TABLE 6
Subjects' Ratings of the Experiment 2

Material	Emotion portrayed by the stimuli	Evaluation scale				
		Aggressive	Happy	Neutral	Sad	Tender
Gravel	Aggressive	**5.67** (2.11-7.57)	3.59 (0-6.26)	0 (0-2.52)	0 (0-0)	0 (0-0)
	Happy	0 (0-0.66)	2.5 (0.69-5.69)	**5.13** (0-8.62)	0 (0-0)	0 (0-0)
	Neutral	0 (0-0)	0 (0-0.04)	**3.05** (0-6.35)	.64 (0-5.54)	1.81 (0-3.51)
	Sad	0 (0-0)	0 (0-0)	0 (0-4.19)	**3.96** (0-6.71)	2.4 (0-6.9)
	Tender	0 (0-0)	0 (0-0)	.51 (0-5.8)	1.74 (0-6.41)	**3.08** (0.53-6.02)
Metal	Aggressive	**2.46** (0-5.6)	.71 (0-5.26)	.03 (0-5.36)	0 (0-0)	0 (0-0)
	Happy	0 (0-1.68)	1.74 (0-5.43)	**3.36** (0-5.96)	0 (0-0)	0 (0-0.62)
	Neutral	0 (0-0)	0 (0-3.18)	**4.71** (0-6.97)	.35 (0-2.72)	.08 (0-4.75)
	Sad	0 (0-0)	0 (0-0)	**2.93** (0-6.88)	1.52 (0-4.82)	2.4 (0-6.55)
	Tender	0 (0-0)	0 (0-0)	.18 (0-4.03)	1.34 (0-3.55)	**4.11** (0-6.75)
Snow	Aggressive	**8.56** (6.46-9.73)	0 (0-0)	0 (0-0)	0 (0-1.61)	0 (0-0)
	Happy	**2.88** (0-6.53)	0 (0-3.11)	1.09 (0-5.89)	0 (0-0.65)	0 (0-0)
	Neutral	0 (0-0.46)	0 (0-0)	1.91 (0-5.88)	**2.29** (0-6.37)	0 (0-3.9)
	Sad	0 (0-0)	0 (0-0)	0 (0-1.85)	**5.62** (0.69-8.07)	2 (0-4.53)
	Tender	0 (0-0)	0 (0-0)	0 (0-2.72)	**6.26** (0.93-7.95)	2.33 (0.68-4.87)
Wood	Aggressive	**5.71** (4.09-7.85)	0 (0-3.2)	0 (0-1.67)	0 (0-0.67)	0 (0-0)
	Happy	0 (0-2.13)	0 (0-0.12)	**6.94** (2.46-9.13)	0 (0-1.32)	0 (0-0)
	Neutral	0 (0-0)	0 (0-0)	**3.52** (0-7.66)	1.77 (0-6.27)	1.94 (0-4.68)
	Sad	0 (0-0)	0 (0-0)	0 (0–5.62)	**4.89** (0.43-7.63)	2.48 (0-5)
	Tender	0 (0-0)	0 (0-0)	2.72 (0-7.07)	**2.59** (0-6.16)	2.58 (0-5.65)

The values [Median (IQR)], grouped for surface material, refer only to trials involving sneakers.

TABLE 7
Correlations Between Listeners' Judgments and the Two Investigated Acoustical Features (H2H and PK), for Each Material and Shoe Type

		Metal		Wood		Gravel		Snow	
		H2H	PK	H2H	PK	H2H	PK	H2H	PK
Sadness	**Dress Shoes**	0.27***	−0.26***	0.35***	−0.33***	0.28***	−0.24***	0.39***	−0.39***
	Sneakers	0.3***	−0.31***	0.3***	−0.27***	0.36***	−0.33***	0.37***	−0.38***
Tenderness	**Dress Shoes**	0.34***	−0.33***	0.39***	−0.39***	0.34***	−0.42***	0.37***	−0.41***
	Sneakers	0.33***	−0.35***	0.38***	−0.4**	0.41***	−0.41***	0.42***	−0.43***
Neutralness	**Dress Shoes**	0.18***	−0.19***	0	0	0.07	−0.14*	0.2***	−0.22***
	Sneakers	0.04	0.01	0	−0.04	0	0	0.02	−0.03
Happiness	**Dress Shoes**	−0.36***	0.34***	−0.39***	0.39***	−0.31***	0.25***	−0.25***	0.26***
	Sneakers	−0.28***	0.28***	−0.32***	0.32***	−0.4***	0.38***	−0.13*	0.14*
Aggressiveness	**Dress Shoes**	−0.43***	0.44***	−0.48***	0.45***	−0.33***	0.49***	−0.56***	0.58***
	Sneakers	−0.45***	0.47***	−0.51***	0.54***	−0.48***	0.48***	−0.62***	0.64***

** represents $p < .05$, ** $p < .01$ and *** $p < .001$.*

value of the aggressive scale (2.69) was less than the happy scale (4.17) although a multiple comparison test showed that this difference was not significant ($p = .954$). For both the types of shoe, the median value of snow was significantly greater than metal, gravel and wood ($p < .001$). The recognition of the happy stimuli was more problematic: they were often rated as neutral, especially the trials with sneakers. However, under the conditions gravel-dress shoes and metal-dress shoes, the happy scale received the greater median (5.25 and 5.03 respectively), although these values were not significantly different from the neutral scale ($p > .289$). For both the types of shoe, the median values of metal and gravel were significantly greater than snow and wood ($p < .001$). As far as the recognition of the neutral stimuli is concerned, for both the types of shoe, the median value of snow was significantly lower than wood, metal and gravel ($p < .001$). The sad stimuli were not recognized when the sounds simulated the metal surface, whereas the subjects' rates on the sad evaluation scale were high with the snow material, both with dress shoes (median = 5, $p < .05$) and sneakers (median = 5.62), although under this last condition no significant difference was found between the sad and tender scales ($p = .07$). For both the types of shoe, the median value of snow was significantly greater than wood, metal and gravel ($p < .001$), while the median value of wood was significantly greater than metal ($p < .001$) Finally, the tender stimuli received the greatest median ratings in the tender scale under the conditions gravel-dress shoes (3.36), wood-dress shoes (3.31), gravel-sneakers (3.08), and metal-sneakers (4.11), although in any case the values were significantly different from the other scales. For dress shoe, the median values of snow was significantly lower than metal, gravel and wood ($p < .001$).

A further statistical analysis was inspired by the works presented in [9] and [4]. First, the relationships between both H2H and PK, and listeners' judgments was measured. For this purpose, 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 7.

To assess whether participants with musical expertise were better at recognizing the simulated emotional walking intentions 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 $\phi < 0.29$, with an average absolute value of 0.12 and SD = 0.07) 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. 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 to assess whether the differences between the ratings of the questionnaire items were significant. No significant main effect was found for both the groups. In addition, a Mann-Whitney-Wilcoxon Test was conducted, for each questionnaire item, on the evaluations of the two groups of participants. No significant differences were found.

Analogously to the first experiment, the identification performances of both the surface materials and shoes were in accordance with the findings reported in previous studies involving the synthesis engine [15], [28].

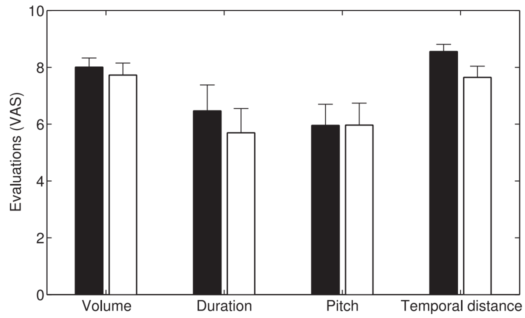

Fig. 5. Graphical representation of the mean and the standard error for participants' answers to questionnaire items Q1, Q2, Q3 and Q4. In black the ratings of the musicians and in white those of the non-musicians.

3.6 Discussion

The results of the second experiment showed that, on average, the emotions portrayed by the synthetic stimuli were correctly identified by participants. Moreover, they allowed to study the influence of the various sound conditions on the identification performances.

Specifically, aggressiveness resulted the emotion best recognized, while sadness and tenderness were confused between each other. The stimuli portraying happiness received ratings lower on the happy scale compared to neutral scale. Such identification performances were in line with results of a similar listening test reported in [17], which involved recordings of synthesized footstep sounds on the same surface materials.

The correlations reported in Table 7 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. 97 percent of these were medium or large according to the guidelines for interpretation of the effects provided by Cohen [36] (as used in [9]), i.e., small ($r \geq .1$), medium ($r \geq .3$), and large ($r \geq .5$), adapted to the Kendall's analysis as indicated in [37], i.e., small ($\tau \geq .065$), medium ($\tau \geq .195$), and large ($\tau \geq .335$). As expected, and consistently with the results reported in [8], the neutral emotion was not associated with medium or large increases or decreases of the two acoustical features considered.

Some tendencies were found regarding the influence of the particular ground material sonically simulated on the recognition of the emotions: for both shoes, snow was the material better associated with the aggressive emotion, while metal was the one worst associated with it; gravel and metal (especially with dress shoes) were better associated with happiness than wood and snow, vice versa for sad; snow with dress shoes was less associated with the tender emotion. These results can in part be explained by considering the spectral content of the sounds. Metal and gravel are sounds with a predominance of high frequency components, while wood and, in particular, snow, have a predominance of low frequency components. In music performance research it has been ascertained that tones with high spectral centroid and/or high pitch 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) [8], [9]. This finding also parallels the results reported in [7] where footstep sounds with spectral content richer of high frequency components caused subjects to feel more aroused and positive.

Consistently with the findings reported in [4] and [17], musical expertise was not associated with a significant increase in the ability to recognize the emotions portrayed by the footstep sounds. The fact that identification performances were similar between the two groups implies that the footstep synthesizer can be involved in applications for a general public. In addition, in both groups no preference for a particular criterium was expressed regarding the first questionnaire item about the subjective criteria which participants relied on to evaluate the expressive content of the auditory stimuli. Regarding both the open comments and the second question about the indicated criteria that participants relied on 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 characterized by slow paces and soft sounds to sad and tender emotions, and fast paces and loud sounds to happy and aggressive emotions. Moreover, more than one participant commented about the difficulty to identify the happy emotion.

4 General Discussion and Conclusion

The first experiment allowed to identify, for each of the five investigated emotions, the mean values and ranges of the two involved acoustical variables. Results showed that participants rendered the various emotions using different sound level and tempo variations similar to those reported in previous studies on real and synthesized footstep sounds [4], [17], as well as on expressive music performance [6], [8].

The second experiment revealed that the algorithms involved for the synthesis of the five expressive walking styles were successful in conveying the emotional information at a level comparable with previous studies [4], [9], [17]. The best performance was achieved for the recognition of the aggressive emotion. Conversely, happiness was the emotion less recognized, and this is likely to be due to the fact that a constant H2H was involved. In a musical context, previous studies showed that happy emotions is often associated to specific rhythmic patterns, such as dotted or syncopated rhythms [38]. A possible rationale is that dotted or syncopated rhythms for happy music may reflect associations between specific actions (e.g., skipping, jumping, dancing) and positive mood. Further experiments are necessary to verify whether the introduction of such a rhythm as a control of the footstep synthesizer is effective in improving the rendering of the happy emotion.

The results of the first experiment showed the presence of an influence of the material sound conditions on the production of aggressive and happy emotions as far as the walking speed is concerned. Faster speeds were associated to solids compared to aggregates for the aggressive emotion, while slower for the happy one. In addition, results showed that subjects were less able to differentiate aggressive from happy emotions on aggregates than on solids. The results of the second experiment showed that the use of certain combinations of surface material and shoe is effective in modulating the emotional rendering. It is possible to provide some guidelines for the use of the involved synthesized sounds: aggressiveness is better rendered using the snow material with both shoe types; gravel and metal (especially with dress shoes) adapts better to happiness than snow and wood; snow adapts better to sad rather than tender; metal adapts better to tender rather than sad. In general, in both production and recognition, the pair happy-aggressive was well differentiated from the pair sad tender. Nevertheless, in both experiments, H2H and PK were not effective in differentiating the two emotions in each pair. Happy and aggressive on the one hand, and sad and tender on the other hand, appeared to be in some cases differentiated by the type of material-shoe combination.

The performances of musically-trained and -untrained subjects had similar trends in both experiments.

Nevertheless, in the first experiment, for each emotion the average values of the two investigated variables were different between the two groups. No plausible explanation was formulated by authors. Further research is needed to understand this result. The most noticeable difference between the two groups was found for the sad emotion. Moreover, musicians appeared to be more able to differentiate sad from tender than non-musicians. Therefore, these results lend support to the MOH only in part. Regarding the differences between the two groups for the sad emotion, several authors (among others [39], [40]) noted that although people generally avoid negative emotional experiences, they often enjoy sadness portrayed in music. Experimental data showed that emotional response to sad music includes also positive emotions such as nostalgia, peacefulness, and wonder.

The results of the first experiment, which involved only imaginary motor activity, are in accordance with those reported in previous studies on real walking [4], [17]. This suggests that expression of emotions in walking is independent from the real or imagined motor activity. This result parallels those of various studies showing that motor imagery results in a somatotopic activation pattern, similar to the same physical movement being executed [19], [20], [21]. Moreover, our findings suggest that the production of emotionally expressive walking sounds in the context of passive sensory motor activity is independent from the anthropomorphic features of the person imagining the simulated walks.

Notably, the capability of participants in producing and identifying emotions in a context of imaginary walking is fully consistent with the findings of various studies that underline the strong link between sound perception and action [23], [24], [25], and their binding with emotion expression [26].

The results of this study also have practical implications for audio applications: they allow for both the design and control of emotionally expressive computerized walking sounds that are more ecologically valid than sounds without performance variations. The reported results can be practically used as guidelines by designers of footstep sounds for videogames or virtual reality in contexts where a control model for artificial walking sound patterns is wanted to convey the emotional state of an avatar. They contribute to the strand of research focusing on the development of more embodied interfaces that take into account emotional expression [10].

Further research is needed to understand how to synthesize walking patterns in such a way that the five investigated emotions can be better distinguished between each other. This is especially relevant for the happy emotion, that was mostly identified as neutral. For this purpose, we plan to investigate variations in the H2H and PK not considered in the present study (e.g., by exploiting results reported in [38]).

Acknowledgments

The research leading to these results has received funding from the Danish Council for Independent Research—Technology and Production Sciences (FTP), grant no. 12-131985.

References

[1] X. F. Li, R. J. Logan, and R. E. Pastore, "Perception of acoustic source characteristics: Walking sounds," *J. Acoust. Soc. Am.*, vol. 90, no. 6, pp. 3036–3049, 1991.

[2] R. Pastore, J. Flint, J. Gaston, and M. Solomon, "Auditory event perception: The source-perception loop for posture in human gait," *Attention, Perception, Psychophys.*, vol. 70, no. 1, pp. 13–29, 2008.

[3] K. Mäkelä, J. Hakulinen, and M. Turunen, "The use of walking sounds in supporting awareness," in *Proc. 9th Int. Conf. Auditory Display*, 2003, pp. 144–147.

[4] B. 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.

[5] 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.

[6] 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.

[7] A. Tajadura-Jiménez, M. Basia, O. Deroy, M. Fairhurst, N. Marquardt, and N. Bianchi-Berthouze, "As light as your footsteps: Altering walking sounds to change perceived body weight, emotional state and gait," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 2943–2952.

[8] 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.

[9] P. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception," *J. Exp. Psychol.: Human Perception Perform.*, vol. 26, no. 6, p. 1797, 2000.

[10] M. Leman, *Embodied Music Cognition and Mediation Technology*. Cambridge, MA, USA: MIT Press, 2008.

[11] P. Cook, "Modeling Bill's Gait: Analysis and parametric synthesis of walking sounds," in *Proc. AES 22nd Int. Conf. Virtual, Synthetic, Entertainment Audio*, 2002, pp. 73–78.

[12] F. Fontana and R. Bresin, "Physics-based sound synthesis and control: Crushing, walking and running by crumpling sounds," in *Proc. Colloq. Musical Informat.*, 2003, pp. 109–114.

[13] A. Farnell, "Marching onwards: Procedural synthetic footsteps for video games and animation," in *Proc. Pure Data Convention*, 2007.

[14] F. Fontana, F. Morreale, T. Regia-Corte, A. Lécuyer, and M. Marchal, "Auditory recognition of floor surfaces by temporal and spectral cues of walking," in *Proc. 17th Int. Conf. Auditory Display*, 2011.

[15] L. Turchet, "Footstep sounds synthesis: Design, implementation, and evaluation of foot-floor interactions, surface materials, shoe types, and walkers' features," *Appl. Acoust.*, in press, 2015.

[16] A. DeWitt and R. Bresin, "Sound design for affective interaction," in *Proc. Int. Conf. Affective Comput. Intell. Interaction*, 2007, vol. 4738, pp. 523–533.

[17] L. Turchet and R. Bresin, "Effects of interactive sonification on emotionally expressive walking styles," *IEEE Trans. Affective Comput.*, vol. 6, pp. 152–164, Apr.-Jun. 2015.

[18] G. Pfurtscheller, R. Leeb, C. Keinrath, D. Friedman, C. Neuper, C. Guger, and M. Slater, "Walking from thought," *Brain Res.*, vol. 1071, no. 1, pp. 145–152, 2006.

[19] J. Decety, M. Jeannerod, and C. Prablanc, "The timing of mentally represented actions," *Behavioural Brain Res.*, vol. 34, no. 1, pp. 35–42, 1989.

[20] C. Porro, M. Francescato, V. Cettolo, M. E. Diamond, P. Baraldi, C. Zuiani, M. Bazzocchi, and P. Di Prampero, "Primary motor and sensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study," *The J. Neurosci.*, vol. 16, no. 23, pp. 7688–7698, 1996.

[21] J. Decety, "The neurophysiological basis of motor imagery," *Behavioural Brain Res.*, vol. 77, no. 1, pp. 45–52, 1996.

[22] W. Young, M. Rodger, and C. Craig, "Perceiving and reenacting spatiotemporal characteristics of walking sounds," *J. Exp. Psychol.: Human Perception Perform.*, vol. 39, no. 2, pp. 464–476, 2013.

[23] E. Kohler, C. Keysers, M. Umiltá, L. Fogassi, V. Gallese, and G. Rizzolatti, "Hearing sounds, understanding actions: Action representation in mirror neurons," *Science*, vol. 297, no. 5582, pp. 846–848, 2002.

[24] S. Aglioti and M. Pazzaglia, "Representing actions through their sound," *Exp. Brain Res.*, vol. 206, no. 2, pp. 141–151, 2010.

[25] A. Lahav, E. Saltzman, and G. Schlaug, "Action representation of sound: Audiomotor recognition network while listening to newly acquired actions," *The J. Neurosci.*, vol. 27, no. 2, pp. 308–314, 2007.

[26] C. Keysers and V. Gazzola, "Expanding the mirror: Vicarious activity for actions, emotions, and sensations," *Current Opinion Neurobiol.*, vol. 19, no. 6, pp. 666–671, 2009.

[27] M. Samson, A. Crowe, P. De Vreede, J. Dessens, S. Duursma, and H. Verhaar, "Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and weight," *Aging Clin. Exp. Res.*, vol. 13, no. 1, pp. 16–21, 2001.

[28] R. Nordahl, S. Serafin, and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," in *Proc. IEEE Virtual Real. Conf.*, 2010, pp. 147–153.

[29] 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.

[30] F. B. Baraldi, G. De Poli, and A. Rodà, "Communicating expressive intentions with a single piano note," *J. New Music Res.*, vol. 35, no. 3, pp. 197–210, 2006.

[31] J. Stevens and J. Hall, "Brightness and loudness as functions of stimulus duration," *Perception Psychophys.*, vol. 1, no. 9, pp. 319–327, 1966.

[32] B. Giordano, Y. Visell, H.-Y. Yao, V. Hayward, J. Cooperstock, and S. McAdams, "Identification of walked-upon materials in auditory, kinesthetic, haptic and audio-haptic conditions," *J. Acoust. Soc. Am.*, vol. 131, pp. 4002–4012, 2012.

[33] L. Turchet and S. Serafin, "Semantic congruence in audio-haptic simulation of footsteps," *Appl. Acoust.*, vol. 75, no. 1, pp. 59–66, 2014.

[34] 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.

[35] F. Konietschke, M. Placzek, F. Schaarschmidt, and L. Hothorn, "nparcomp: An r software package for nonparametric multiple comparisons and simultaneous confidence intervals," *J. Stat. Softw.*, vol. 61, pp. 1–17, 2014.

[36] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd revised ed. New York, NY, USA: Academic, 1988.

[37] 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. Psychol. Meas.*, vol. 53, no. 1, pp. 87–92, 1993.

[38] M. Adachi and S. E. Trehub, "Canadian and Japanese preschoolers' creation of happy and sad songs," *Psychomusicology: Music, Mind Brain*, vol. 21, no. 1, pp. 69–82, 2011.

[39] J. K. Vuoskoski, W. F. Thompson, D. McIlwain, and T. Eerola, "Who enjoys listening to sad music and why?" *Music Perception: An Interdisciplinary J.*, vol. 29, no. 3, pp. 311–317, 2012.

[40] A. Kawakami, K. Furukawa, K. Katahira, and K. Okanoya, "Sad music induces pleasant emotion," *Frontiers Psychol.*, vol. 4, no. 311, pp. 1–15, 2013.

Luca Turchet received the master degrees in computer science in 2006 from the University of Verona, in classical guitar and composition in 2007 and 2009, respectively, from Music Conservatory of Verona, and in electronic music in 2015 from the Royal College of Music of Stockholm. 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 are in sonic interaction design, multimodal systems, human-computer interaction, and perception.

Antonio Rodà received the master's degree in electronic engineering from the University of Padova, Italy in 1996, and the PhD degree in audiovisual studies from the University of Udine, Italy in 2007. Since 1997, he is a member of Centro di Sonologia Computazionale (CSC), University of Padova. He is an author and coauthor of more the 80 paper on national and international journals and peer reviewed conferences. He is currently an assistant research professor at the Department of Information Engineering, University of Padova.