Extended scientific report for the ConGAS Short Term Scientific Mission:

Source perception in walking sounds - walkers' emotion, gender, weight and shoes properties.

Bruno L. Giordano[*]

Contents

---

[*]Electronic mail: bruno.giordano@unipd.it

1 Introduction

A relevant issue to the study of music performance and perception, concerns the ability of players to express emotional intentions and the ability of listeners to recognize them. Several works studied encoding and perception of emotional intentions with musical instruments (eg. Gabrielsson and Juslin, 1996; Juslin, 1997a, 1997b; Bresin and Friberg, 2000; see Juslin and Laukka, 2003 for a review). Typically the same score was played with different emotional intensions, and listeners recognition capabilities were assessed via either categorization, or rating tasks. Using this methodology, properties of the musical structure associated with different emotional intentions were highlighted, both for performance and perception. Clynes (1977, cit. in Gabrielsson & Juslin, 1996) argued that expression of emotions should be based on structures independent of modalities, i.e. should present similar traits independently of whether the medium for their expression is music, voice, facial expression, or a gesture. Consistently, independently of whether an emotional intention is expressed through instrumental music, singing voice, or speech, equal emotions are associated with similar variations in properties such as average tempo, articulation, or loudness (Juslin & Laukka, 2003). The strong similarities in emotions expression highlighted between speech and music performance, support furthermore Clynes' 1997 hypothesis. Thus one is motivated in expecting that structures similar to those highlighted in speech, and music, might be used to express emotions with one of the most frequently executed gesture: walking. That walking could be a good vehicle for the expression of emotions is supported by at least one study, which investigated the acoustical structure of this acoustical event outlining features known to differentiate between emotional playing styles in music (Bresin & Dahl, 2003). In this study several similarities between articulation in piano playing and walking or running were drawn, where the amount of temporal overlap among adjacent tones/footstep sounds distinguished between staccato and legato articulation. As differences in articulation have already been found to be used in musical expression of emotions (eg. Bresin & Battel, 2000), one is supported in expecting an analogue relevance in the walking performance. A final relevant point is in favor of the study of emotions expression through walking. For obvious reasons the above mentioned studies investigated musical rendering of emotions using performers trained in music, i.e., performers that received implicit and explicit training concerning expressive rendering of performances. This requirement is instead not necessary if expression of emotions is investigated with walking, where musically naïve performers can be used as well. Thus, the case of walking provides a valuable occasion to test for the generality of the rules used in musical conditions for emotions expression, i.e. whether they are found outside the concert hall too. To summarize, the first goal of this study has been to validate the modality independence hypothesis, and to test for the generalizability of musical performance rules to an everyday gesture: walking.

Another parallel interest of this study concerned the perceptual level, i.e. on the recognition of the emotional intentions of the walkers conceived as properties of the sound source. Several previous studies outlined a fine tuning of the auditory system to the properties of the sound source, as resulting from sometimes surprisingly good recognition performance of attributes such as the shape of a struck plate (Kunkler-Peck & Turvey, 2000) or the integrity of an object bouncing or shattering on the floor (Warren & Verbrugge, 1984) (see McAdams, 1993 for an introduction to source recognition; Giordano, 2005 for an updated presentation of the literature in this field). Given a plausible influence of the emotion of the walker on the structure of walking sounds (see above), it is likely that our auditory system developed mood-recognition capabilities based on these types of signals. This ability would serve a general purpose of the perceptual representation: predicting the environment, in this case the actions of the sound source on the basis of the incoming signal (no doubts that we are likely react differently depending on whether we recognize that an approaching walker is extremely angry and thus potentially dangerous rather than cheerful). Two previous studies investigated auditory recognition of the properties of the walking sound source. In a first study Li, Logan, and Pastore (1991) studied perception of the gender of the walker, finding extremely high recognition performances. A second study (Pastore, Flint, Gaston, & Solomon, submitted) investigated recognition of the posture of the walker, i.e. discrimination between an upright and stopped posture. Even in presence of correctness-feedback, recognition performance was found to be impaired with respect to gender recovery. Such a difference might plausibly hypothesized to be due to the relative everyday-life-usefulness of recognizing the gender, rather than the posture of a walker. Such a question, however, remains unanswered at present, given the absence of studies that carried an explicit comparison of the recognition performance for different source attributes and, in the specific, for different attributes of the walking sound source (indeed, why a perceiver would develop perhaps computationally demanding perceptual strategies if they would not generate adaptively useful contents?). This has been done in the current study, where listener have been asked to recognize several different attributes of the walkers, mood included, allowing to draw a hierarchy of source attributes based on recognition performance. In such a way we aimed also at providing an answer to the question concerning which source attributes are relevant to the walking sounds perceiver in everyday conditions.

The methodology adopted in the present study has been outlined by Li et al. (1991), and labeled lately as the study of the “source-perception loop”. Accordingly, the study was divided in three phases. Firstly, the mapping of the source properties to the structure of the acoustical signal was highlighted, investigating the acoustical properties most strongly influenced by the parameters of the walking sound source, mood included. This stage required carrying a walking performance experiment, and thus provided the necessary data for the validation of the modality-independence hypothesis for emotions expression (see above). Secondly, the source properties recognition capabilities were investigated in a listening test. Finally, the same perceptual data were explained in terms of acoustical descriptors, thus providing information concerning the acoustical basis for the perception of the walking sound source.

2 Walking experiment

2.1 Methods

2.1.1 Procedure

Participants were asked to walk along a 10 m track marked on a linoleum floor with scotch tape. They were instructed to walk as if either they were feeling a certain emotion, or as if they were in a normal, i.e. emotionless mood. Four basic emotions were investigated: happiness, sadness, anger and fear. Before the beginning of the recording session participants practiced each of the walking styles three times. Five trials were recorded for each of the walking styles. Walking styles were performed in blocked-randomized order, i.e., all five walking styles were performed in random order before the same one was repeated.

Sounds were recorded with a Brüel & Kjær DPA type 4021 microphone, connected to an USB Tascam US-122 soundcard (44100-Hz sampling rate, 16-bit resolution). The microphone was placed on the side of the walking path, at half of its length and 1 m above ground.

2.1.2 Participants

Seven walkers took part in the experiment on a voluntary basis (age: 62 - 28; 2 females, 5 males). All participants had no prior training with either singing or a musical instrument. This choice was made in order to avoid participants generalizing to walking performance rules for emotional expression learnt during musical training. Participants were asked to wear a pair of their one shoes capable of generating rather loud sounds. For one of the participants the shoe soles were not hard enough to generate reasonably loud sounds. In order to increase the loudness of the walking sounds, several metallic pins were inserted under the soles. Table 1 reports anthropometric measures, age, gender and shoes properties for each of the walkers. Sole hardness was estimated roughly along an ordinal scale.

2.1.3 Acoustical analyses

Walking sounds comprised a variable number of footstep sounds. Typically, a footstep sound should be given by two consequent impact sounds, generated by a heel strike followed by a toe strike (see Figure 1). This pattern, however, was not observed frequently in the recordings, as more typically only one impact sound per footstep was generated, given either by the absence of temporal separation between the heel and toe strikes, or by the absence of appreciable acoustical energy generated by the heel strike. It should also be noted that, as the walkers in the experiment were not asked to wear tight clothes, additional acoustical signals were occasionally generated by the friction of the clothes.

Given the positioning of the microphone in the center of the walking track, the amplitude of the footstep sounds markedly tended to increase from the beginning to the middle of the walking sequence and to decrease from the middle to the end. The same pattern characterized the walking rhythm, the initial

Table 1: Anthropometric measures, age, gender, shoe size and hardness for the tested walkers. M = male; F = female; L. = leather; W. = wood; R. = rubber; S. = soft; H. = hard.

Walker	O.	C.	M.	S.	H.	B.	L.
Age	32	62	28	41	34	36	32
Gender	M	F	M	M	M	M	F
Weight (kg)	85.5	70	73	100	72	92	58
Height (cm)	192	172	179	192	189	167	181
Ground-hip (cm)	111	107	100	109	114	96	108
Ankle-knee (cm)	54	49	49	49	47	45	49
Knee-hip (cm)	48	59	53	51	62	56	54
Shoe length (cm)	32	28	29.5	28	29.5	30	27
Shoe width (cm)	15	8.5	11.5	10	11.5	11.5	9.5
Sole material	L.+R.	W.+R.	H.R.	W.+R.	H.R.	S.R.	W.+R.
Sole hardness	2	4	3	4	3	1	4

and final footsteps tending to be more separated temporally than the central ones. In order to eliminate the influence of this latter effect on the extracted rhythm/tempo measures, it was decided to limit acoustical analyses to the central portion of the recorded signals sounds, where gait adjustments associated to the initiation and termination of the walking were not present or weak.

2.1.3.1 Extraction of walking excerpts The seven middle footstep sounds of the walking sequences $x(t)$ were extracted, applying a criterion based on the analysis of the amplitude envelope $E(t)$, defined as:

$$E(t) = |x(t) + i\mathrm{H}[x(t)]| \tag{1}$$

where H is the Hilbert transform of $x(t)$ (Hartmann, 1997). Further, $E(t)$ was forward-reverse filtered with a third-order Butterworth filter with a low-pass cutoff frequency of 16 Hz (McAdams, Chaigne, & Roussarie, 2004), thus limiting the influence of hearing-range amplitude fluctuations on the extracted envelope.

Local $E(t)$ peaks were then isolated. Only those whose amplitude was at least 10 dB higher than the average envelope amplitude of the walking sequence were retained. Then, assuming a minimum temporal distance of 250 ms between adjacent footstep sounds (Nilsson & Thorstensson, 1987, 1989), one amplitude peak per footstep sound was extracted, i.e. the loudest. Finally, seven footstep were isolated, i.e. that characterized by the highest amplitude peak, and thus generated in the position closest to the microphone, plus the three succeeding and preceding ones. For those sequences for which less than seven footsteps were found the entire signal was kept; for those sequences were less than three

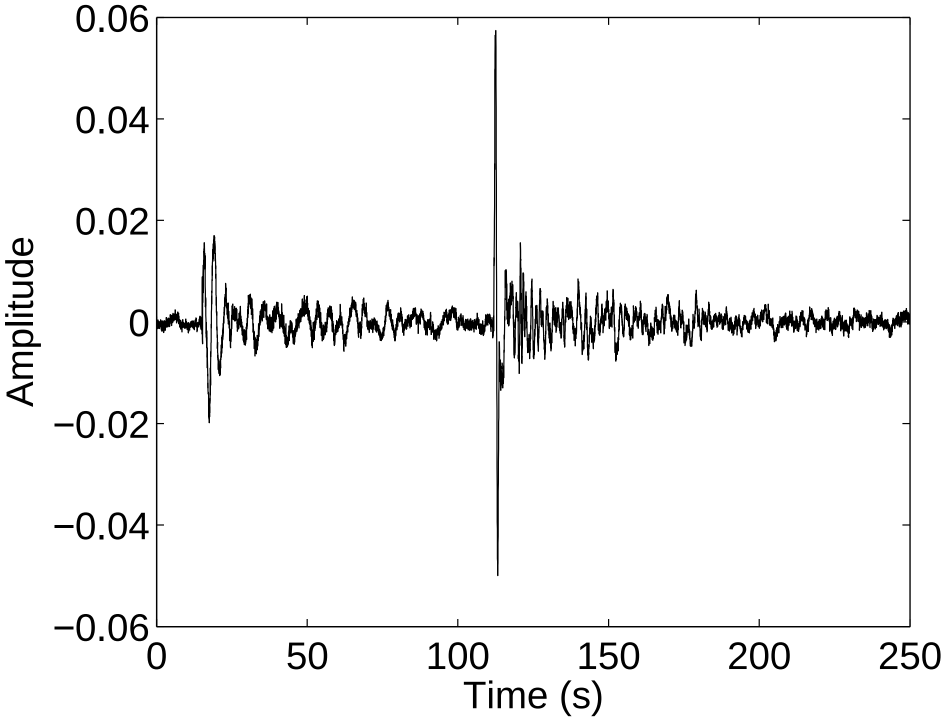

Figure 1: Example of footstep sound given by a heel strike followed by a toe strike.

footsteps succeeding/preceding the loudest one were present, seven footsteps were extracted considering more than three of the footsteps preceding/succeding the loudest.

For each of the walking excerpts four acoustical descriptors were extracted: the average temporal distance among adjacent footstep peaks (IOI_{mea}) and its standard deviation (IOI_{std}) and the average envelope amplitude of the footstep peaks (Amp_{mea}) and its standard deviation (Amp_{std}). Selection of the stimuli presented in the listening test was based on these descriptors (see Section 4).

2.1.3.2 Acoustical characterization of walking excerpts Analyses focused mainly on the amplitude envelope $E(t)$, computed as described in Section 2.1.3.1. The first phase of the analysis process was meant to locate the local peaks of the amplitude envelope resulting from the shoe sole strikes, assumed to represent those portions of the walking excerpts with the highest perceptual salience.

First, those envelope portions where amplitude was lower than a given threshold were discarded as perceptually irrelevant. The threshold was established on the basis of the analysis of the background noise present in the recordings. Five 500 ms background noise samples were thus extracted from the recordings of each walker. The average envelope amplitude was then calculated, and the threshold was set at this value plus 10 dB.

Amplitude peaks were then isolated. Peaks associated to above-threshold envelope portions shorter than 20 ms were discarded (it might be of some value to observe that Pastore et al., submitted, found shoe sole impact sounds not to be shorter than 40 ms).

Low-amplitude above-threshold envelope portions, resulting from the friction between the clothes of the walkers, were occasionally present and associated with local amplitude peaks. They were assumed having limited relevance in determining perceptual estimates of walkers' and shoes' parameters (for example all walkers wore pants, so that frictions of pants clothes could not reliably discriminate between walkers' gender). Clothes friction-related envelope peaks were thus isolated from the shoe sole-related ones, on the basis of the fact that friction sounds are characterized by a slower onset than the impact shoe sole strike sounds. Thus, for each of the extracted peaks the preceding 20 ms of the envelope were used to calculate the level increase rate, by means of linear regression. Peaks associated to a level rise of less than 100 dB/s were assumed to identify friction sounds and were consequently discarded.

The following assumptions were finally applied to identify shoe sole strike peaks, and to eventually discriminate between heel and toe strike peaks:

- the minimum temporal distance between temporally adjacent footsteps is 250 ms (Nilsson & Thorstensson, 1987, 1989);
- the heel strike always precedes the toe strike;
- heel and toe strike envelope peaks are temporally adjacent;
- the sum of the heel and toe strikes peak envelope amplitudes is higher than the sum of any of these with peak amplitudes not related to shoe sole strike sounds;
- the maximum temporal distance between heel and toe strikes is 150 ms (Pastore et al., submitted, finds an average heel-to-toe strikes distance of 112 ms).

An example of the results of the peak extraction procedure is shown in Figure 2.

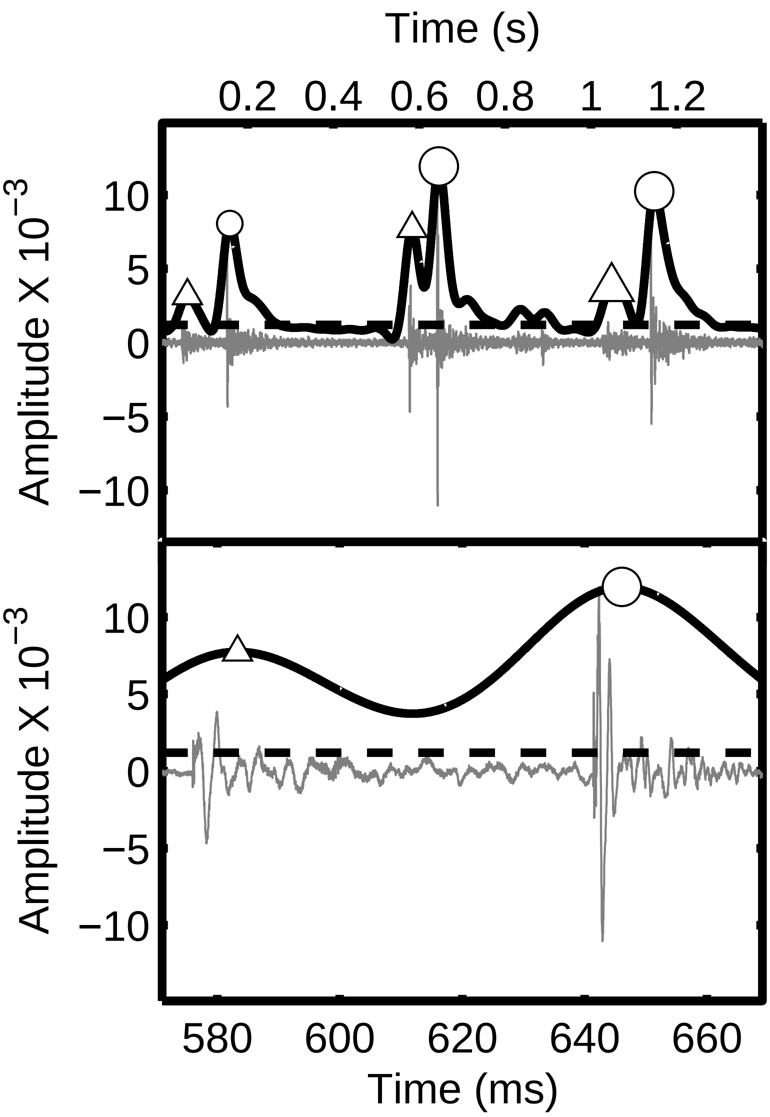

Figure 2: Extraction of heel and toe strike amplitude envelope peaks. The waveform excerpt (grey) is shown along with the amplitude envelope (black), and with the heel and toe strike peaks (white triangles and circles, respectively). The dashed line shows the amplitude threshold used to discard perceptually irrelevant peaks. For convenience the peak waveform amplitude is normalized to the peak envelope amplitude. The detail in the bottom panel shows the lag of the envelope peaks with respect to the impact sounds onset.

Two aspects of the results of this analysis should be pointed out. Firstly, the threshold used for peak detection at this stage was, for the vast majority of the analyzed excerpts, lower than that used to extract them from the original recordings (cf., Section 2.1.3.1). This was due to the fact that the threshold used in the first stage depended on the average envelope amplitude of the recorded signal, louder recordings having higher thresholds. Consequently, the number of footsteps extracted in this latter analysis stage was frequently higher than that resulting from the excerpt extraction stage, and thus slightly variable across signals (average: 7.07 footsteps per walking excerpt). Secondly, in not all cases the footstep sound comprised two distinct impact sounds (the heel and toe strike sounds). Consequently, in this case it was assumed that heel and toe strikes were both present and contemporaneous. It should finally be noted that the heel and toe strike envelope peaks lag the onset of the related impact sounds of approximately 10 ms (see Figure 2, bottom panel).

For each walking excerpt several acoustical descriptors were extracted. First, the amplitude of the toe strike and heel strike envelope peaks ($T.amp$; $H.amp$), measured in dB from the envelope analysis threshold (dB_{thr}). Second, the temporal distance between adjacent toe strikes ($T2T$) and between heel strikes and the succeeding toe strikes ($H2T$), calculated taking into account the temporal location of the envelope peaks. Third, the temporal distance between the offset of toe strikes and the onset of the succeeding heel strikes ($T2H$), defined as the temporal position of the last above threshold toe strike sample and as the temporal location of the first above threshold heel strike envelope sample, respectively. Fourth, the probability of observing a heel strike temporally separated from the toe strike $p(h)$. Fifth, a set of spectral measures for each toe and heel strike, derived from the fast Fourier transform of the first 2048 samples of the impact sound (Hanning window), the position of the first analysis sample corresponding to the temporal location of the envelope peak minus 10 ms. Thus, for both toe and heel strike impact sounds the spectral center of gravity, the amplitude weighted average of frequency, was calculated ($T.SCG$ and $H.SCG$ respectively), considering frequency bins between 16 and 16000 Hz. Also, for both the toe and heel strike impact sounds the spectral mode was extracted ($T.mod$ and $H.mod$ respectively), i.e., the frequency of the bin characterized by the highest amplitude, calculated taking into account only the 16-16000 Hz region. Sixth, the duration Dur was computed, defined as the temporal distance between the first heel strike and the last toe strike in the walking excerpt. Given its variability within the walking excerpts (see above), also the number of footsteps per was extracted (Nf).

From all these indices but $p(h)$, Dur and Nf, two measures were computed in order to characterize the entire excerpt, i.e. the average and the standard deviation of the measures within the analyzed signal. For each of these indices the average and standard deviation measures are indicated, respectively, by the *mea* and *std* subscripts (e.g. $T.amp_{mea}$ is the average of the toe strike amplitudes within the walking excerpt; $H.mod_{std}$ is the standard deviation of the spectral modes of the heel impact sounds within the walking excerpt).

2.2 Results

2.2.1 Acoustical correlates of walker parameters

For each of the measured parameters of the walking sound source but walkers' age, acoustical correlates were preliminary investigated by means of univariate ANOVA models. The purpose of this analysis was to highlight the acoustical descriptor that differentiated better among the levels of each of the source parameters. Thus, for each of the source parameters/acoustical descriptor pairs, a univariate ANOVA model was computed, with the walker parameter as independent variable, and the acoustical descriptor as dependent variable. For each of these models, a measure of the size of the effect of the independent variable was computed, the partial η^2 statistics (Cohen, 1973), where the higher its value the higher the influence of the independent on the dependent. The results of this analysis are shown in Table 2. Figures 3-5 show, for each of the source parameters, the maximally associated acoustical descriptors, i.e. that for which the ANOVA model yielded the highest partial η^2.

Table 2: Analysis of the acoustical correlates of the walking sound source parameters. For each of the source/acoustical parameters pairs the partial η^2 measure of effect size is shown. For the models for which the effect resulted significant ($\alpha = 0.05$) the partial η^2 measure is shown in bold face.

	Gender	Weight	Height	Ground-Hip	Ankle-Knee	Knee-Hip	Shoe length	Shoe width	Sole hardness	Emotion
$T.amp_{mea}$	**0.071**	**0.506**	**0.505**	**0.506**	0.041	**0.506**	**0.212**	**0.485**	0.040	**0.326**
$H.amp_{mea}$	**0.049**	**0.550**	**0.497**	**0.550**	**0.049**	**0.550**	**0.136**	**0.504**	**0.056**	**0.247**
$T2T_{mea}$	**0.042**	**0.158**	**0.158**	**0.158**	0.015	**0.158**	**0.157**	**0.157**	0.020	**0.124**
$H2T_{mea}$	**0.152**	**0.624**	**0.554**	**0.624**	**0.094**	**0.624**	**0.293**	**0.620**	**0.127**	0.012
$T.SCG_{mea}$	0.020	**0.804**	**0.727**	**0.804**	**0.454**	**0.804**	**0.624**	**0.437**	**0.470**	**0.061**
$H.SCG_{mea}$	**0.028**	**0.842**	**0.528**	**0.842**	**0.348**	**0.842**	**0.446**	**0.543**	**0.424**	0.030
$T.mod_{mea}$	**0.192**	**0.215**	**0.199**	**0.215**	**0.100**	**0.215**	**0.186**	**0.214**	**0.155**	0.032
$H.mod_{mea}$	**0.069**	**0.173**	**0.154**	**0.173**	**0.079**	**0.173**	**0.173**	**0.129**	**0.119**	0.037
$T2H_{mea}$	**0.040**	**0.185**	**0.185**	**0.185**	0.019	**0.185**	**0.181**	**0.183**	0.021	**0.149**
$T.amp_{std}$	**0.048**	**0.309**	**0.120**	**0.309**	**0.290**	**0.309**	**0.251**	**0.263**	**0.245**	0.017
$H.amp_{std}$	0.002	**0.303**	**0.201**	**0.303**	**0.128**	**0.303**	**0.206**	**0.296**	**0.131**	0.030
$T2T_{std}$	0.016	**0.094**	**0.089**	**0.094**	0.015	**0.094**	**0.091**	**0.092**	0.016	**0.107**
$H2T_{std}$	**0.142**	**0.235**	**0.235**	**0.235**	**0.071**	**0.235**	**0.171**	**0.229**	**0.071**	0.038
$T.SCG_{std}$	0.005	**0.153**	**0.080**	**0.153**	**0.092**	**0.153**	**0.100**	**0.091**	**0.100**	0.024
$H.SCG_{std}$	0.005	**0.155**	**0.128**	**0.155**	**0.084**	**0.155**	**0.133**	**0.104**	**0.117**	0.018
$T.mod_{std}$	0.014	0.041	0.038	0.041	0.018	0.041	0.040	0.0190	0.036	0.044
$H.mod_{std}$	0.011	**0.124**	**0.102**	**0.124**	**0.063**	**0.124**	**0.113**	**0.064**	**0.092**	0.037
$T2H_{std}$	**0.024**	**0.091**	**0.088**	**0.091**	0.014	**0.091**	**0.089**	**0.090**	0.020	**0.105**
$p(h)$	**0.157**	**0.651**	**0.584**	**0.651**	**0.106**	**0.651**	**0.321**	**0.650**	**0.140**	0.006
Dur	0.010	**0.226**	**0.214**	**0.226**	0.015	**0.226**	**0.214**	**0.216**	0.014	**0.274**
Nf	**0.043**	**0.169**	**0.127**	**0.169**	**0.101**	**0.169**	**0.160**	**0.154**	**0.098**	**0.075**

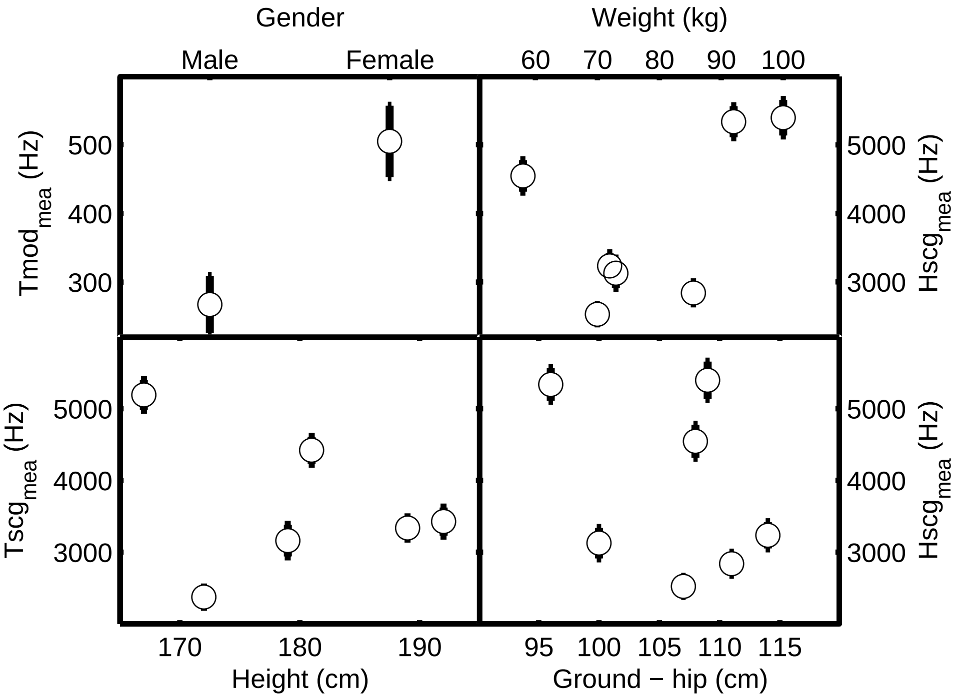

Figure 3: Best acoustical correlates for walker gender, weight and height and for the ground-hip distance. White circles: average values. Error bars bracket 95% confidence intervals for the average value.

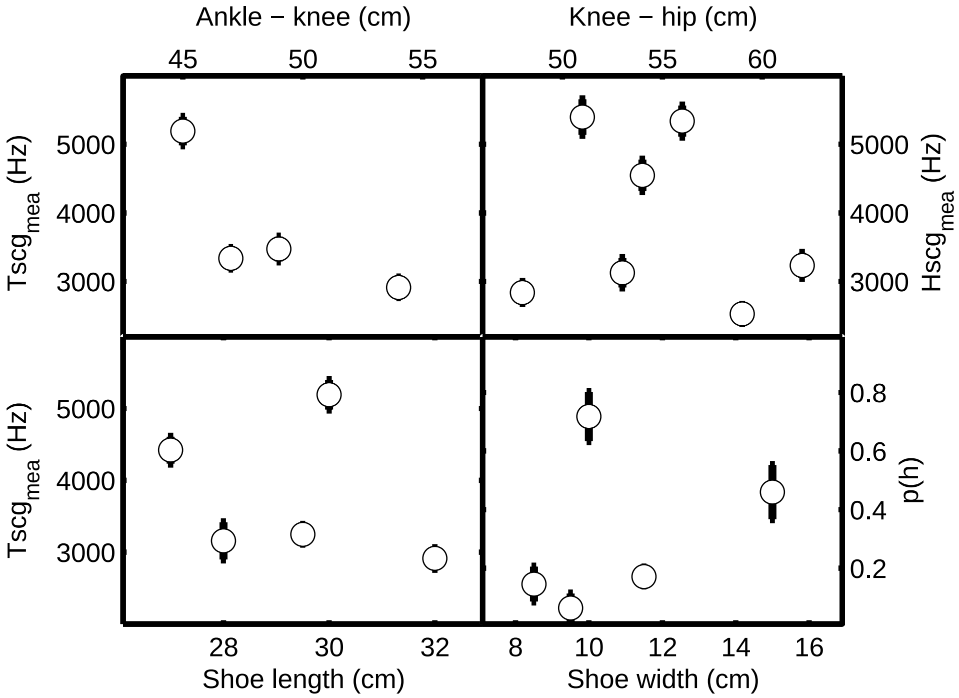

Figure 4: Best acoustical correlates for ankle-knee and knee-hip distance and for show length and width. White circles: average values. Error bars bracket 95% confidence intervals for the average value.

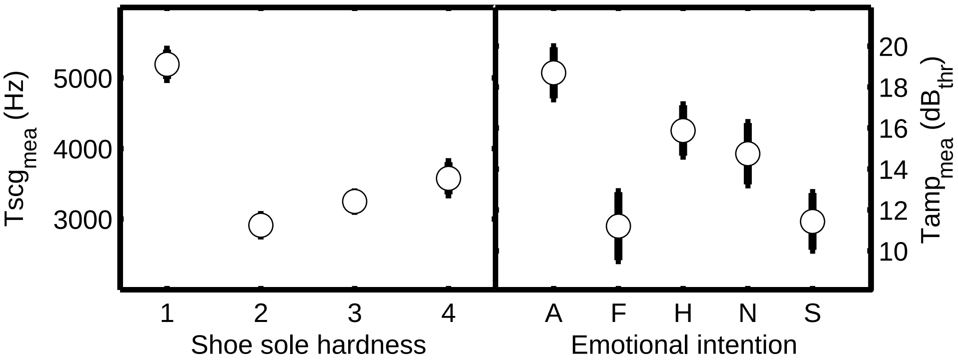

Figure 5: Best acoustical correlates for shoe sole hardness and of the emotional intention (A = angry; F = fearful; H = happy; N = normal; S = sad). White circles: average values. Error bars bracket 95% confidence intervals for the average value.

2.2.2 Acoustical correlates of emotional intention

The same analysis carried to investigate the acoustical correlates of walker parameters and of emotional walking styles in general was applied to the study of the acoustical correlates of the different emotional intentions separately. Univariate ANOVA models were thus used, with the acoustical descriptors as dependent variables. In order to consider each emotional intention separately, this latter coded as a binary variable, being 1 for those excerpts generated with a given emotional intention and 0 for those excerpts generated with any other emotional intention. The results of this analysis is shown in Table .

Table 3: Analysis of the acoustical correlates of emotional walking intentions. For each of the emotional intention/acoustical parameter pairs the partial η^2 measure of effect size is shown. For the models for which the effect resulted significant ($\alpha = 0.05$) the partial η^2 measure is shown in bold face.

	Anger	Fear	Happiness	Normal	Sadness
$T.amp_{mea}$	**0.190**	**0.104**	**0.023**	0.001	**0.326**
$H.amp_{mea}$	**0.135**	**0.107**	0.021	0.001	**0.247**
$T2T_{mea}$	0.019	**0.108**	0.017	0.010	**0.124**
$H2T_{mea}$	0.004	0.003	0.000	0.007	0.002
$T.SCG_{mea}$	**0.033**	0.022	0.004	0.001	**0.061**
$H.SCG_{mea}$	**0.024**	0.011	0.000	0.000	0.002
$T.mod_{mea}$	0.007	0.000	0.000	0.005	0.028
$H.mod_{mea}$	**0.025**	0.003	0.003	0.002	0.012
$T2H_{mea}$	**0.032**	**0.120**	0.019	0.010	**0.149**
$T.amp_{std}$	0.002	0.001	0.017	0.001	0.001
$H.amp_{std}$	0.001	0.002	0.012	0.002	0.021
$T2T_{std}$	0.018	**0.095**	0.008	0.011	**0.107**
$H2T_{std}$	0.009	0.005	0.005	0.021	0.008
$T.SCG_{std}$	0.000	0.012	0.007	0.008	0.003
$H.SCG_{std}$	0.005	0.004	0.001	0.009	0.004
$T.mod_{std}$	0.006	0.000	0.002	0.007	0.040
$H.mod_{std}$	**0.022**	0.001	0.005	0.004	0.014
$T2H_{std}$	0.010	**0.100**	0.011	0.010	**0.105**
$p(h)$	0.000	0.005	0.001	0.003	0.000
Dur	**0.078**	**0.149**	**0.044**	0.017	**0.274**
Nf	0.001	**0.040**	**0.033**	0.007	**0.075**

As can be seen, out of the 21 investigated acoustical descriptors only 12 discriminated significantly between at least one of the emotional styles ($T.amp_{mea}$, $H.amp_{mea}$,$T2T_{mea}$, $T.SCG_{mea}$, $H.SCG_{mea}$, $H.mod_{mea}$, $T2H_{mea}$, $T2T_{std}$, $H.mod_{std}$, $T2H_{std}$, Dur, Nf). Interestingly, the majority of the non-significantly discriminating variables modeled the temporal variability of the signal and, in particular, the variability of the spectral properties of the footstep sounds. Figures 6-8 show, for each of the emotional walking styles, the average value of each of these 12 descriptors.

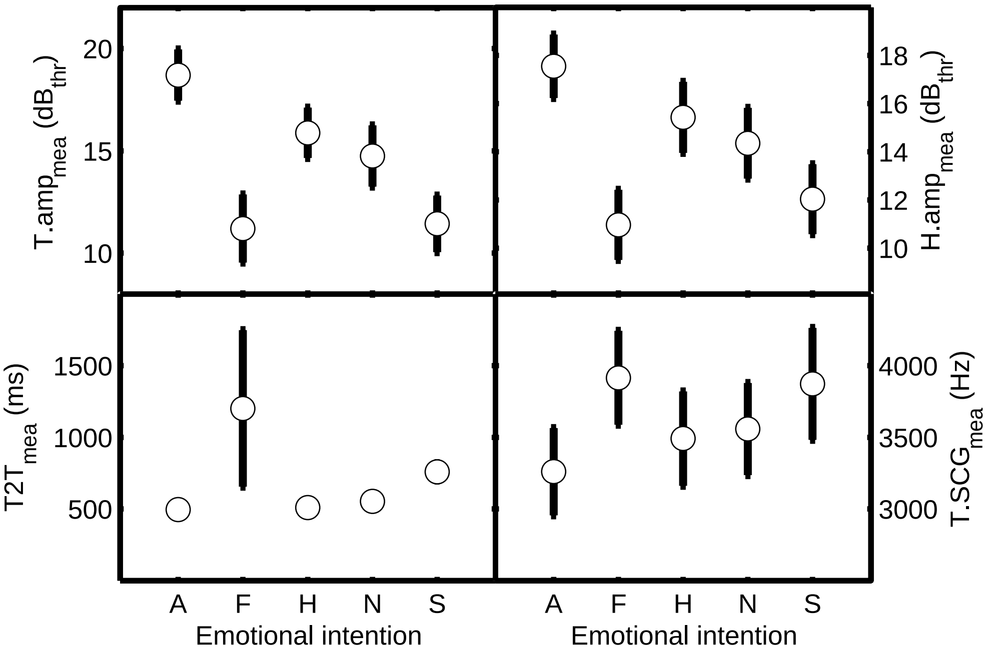

Figure 6: Acoustical correlates of emotional walking intention. White circles: average values. Error bars bracket 95% confidence intervals for the average value.

The consistency between emotions expression in walking and in music performance was finally tested. Table 4 contrasts the acoustical properties of musical performances when one of the four emotions investigated in the present study are expressed (adapted after Juslin, 2001). Only those acoustical features which could be measured with the descriptors presented in Section 2.1.3.2 are

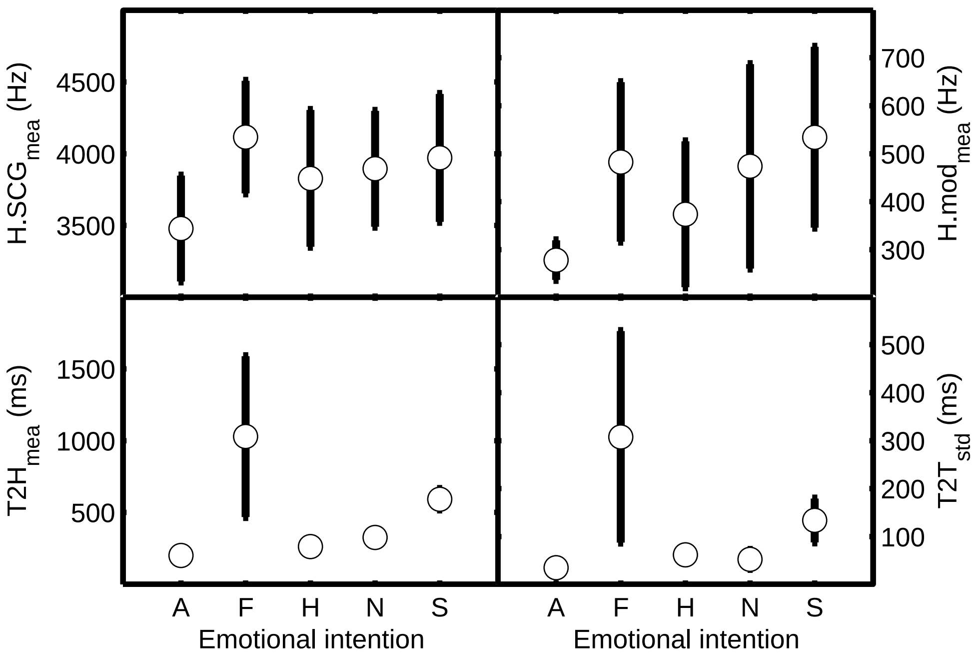

Figure 7: Acoustical correlates of emotional walking intention. White circles: average values. Error bars bracket 95% confidence intervals for the average value.

included. The acoustical correlates of emotional expression in music performance were converted in specific hypotheses concerning the acoustical properties of walking performances, as shown in Table 5. For each of the considered walking sounds descriptors these hypotheses were evaluated through multiple t-tests, which tested for the presence of significant differences between the considered emotional walking intentions with respect to the considered acoustical descriptor. For each of the t-tests the critical p-value for the rejection of the null hypothesis of no difference between emotional intentions was fixed at 0.05. Table 5 also outlines the results of this test.

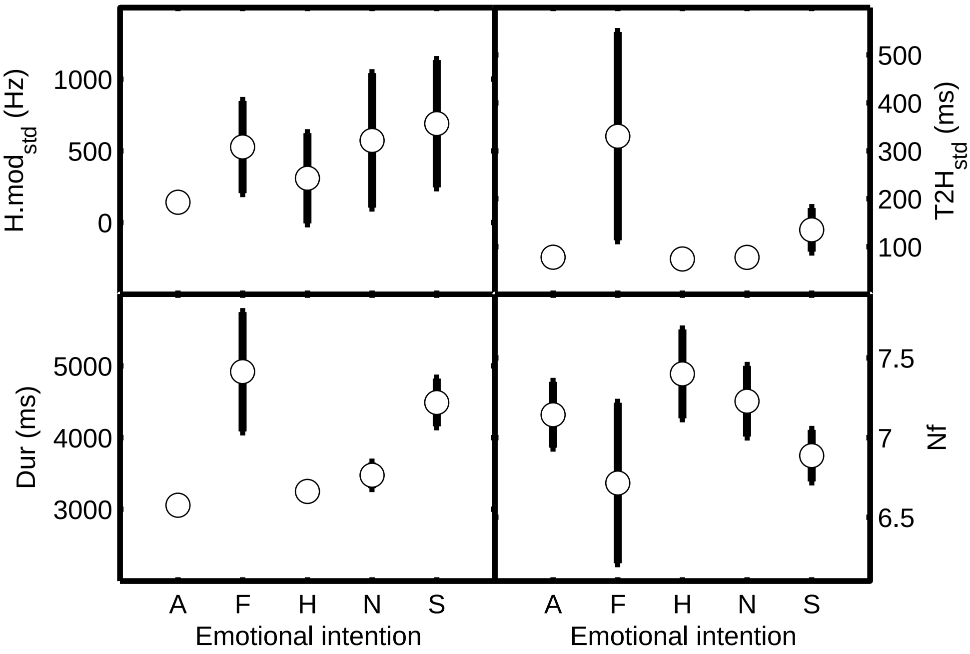

Figure 8: Acoustical correlates of emotional walking intention. White circles: average values. Error bars bracket 95% confidence intervals for the average value.

Acoustical property	**Anger**	**Fear**	**Happiness**	**Sadness**
Mean tempo	fast	fast	fast	slow
Tempo variability	small	large	small	
Sound level	high	very low	high	low
Articulation	staccato	staccato	staccato	legato
Timbre	sharp	soft	bright	dull
Articulation variability			large	small

Table 4: Musicians' use of acoustic cues when communicating emotion in music performance.

Acoustical property	**Tested Descriptor**	**Music performance**	**Walking performance**
Mean tempo	$T2T_{mea}$	$(A. = F. = H.) < S.$	$(H. = A.) < (S. = F.)$
Tempo variability	$T2T_{std}$	$(A. = H.) < F.$	$A. < H. < F.$
Sound level	$T.amp_{mea}$	$F. < S. < (A. = H.)$	$(S. = F.) < H. < A.$
	$H.amp_{mea}$	$F. < S. < (A. = H.)$	$(S. = F.) < H. < A.$
Articulation	$T2H_{mea}$	$S. < (A. = F. = H.).$	$A. < H. < (S. = F.)$
Timbre	$T.SCG_{mea}$	$(F. = S.) < (A. = H.)$	$(H. = A.) < (S. = F.)$ with $H. = S. = F.$
	$H.SCG_{mea}$	$(F. = S.) < (A. = H.)$	$H. = A. = S. = F.$ with $A. < F.$
Articulation variability	$T2H_{std}$	$S. < H.$	$H. < S.$

Table 5: Comparison of emotions expression in music performance and in walking performance.

As can be seen strong similarities are found between the acoustical correlates for emotions expression in music and walking for the tempo- and sound level-related descriptors, while strong differences are found for the articulation and timbre-related descriptors. Concerning both tempo-related and loudness-related descriptors, almost exactly the same ordering among emotional intentions which characterizes music performance is found in walking performance. Concerning articulation, the orderings highlighted among emotional intentions in walking are almost the perfect contrary of those outlined for walking performance. There is thus the doubt that the adopted measure of articulation $T2H$ reflects more differences in tempo, rather than articulation itself. This could be due to the fact that while care might be used by the walker to vary tempo and loudness according to the different emotional intentions, the length of the footstep sounds would remain uncontrolled (and probably would be rather hard to control), being kept constant across emotional intentions. Gross differences are found also concerning timbre related attributes, where, contrary to music performance, the "sad" and "fearful" performances are characterized by a lower spectral centra of gravity, and thus by a brighter timbre.

3 Discussion

Properties of the acoustical structure influenced by walkers parameters such as gender or shoe sole hardness and by the emotional intentions of the walker were sought and highlighted. Consistently with the hypotheses outlined in the Section 1, several strong similarities were found between musical and walking expression of emotions. In particular almost the same ordering among emotions were found in the two domains for the average tempo, for tempo variability and for average signal level. Substantial differences were instead found with articulation-related descriptors. In particular with walking anger and happiness were characterized by a more legato-like articulation, while sadness and fear were characterized by a more staccato-like articulation style. Thus, contrary results to those observed in the music performance field in general, and in the piano performance field in particular (Bresin & Dahl, 2003) were observed. Strong differences between the walking and music domains were also observed with respect to timbre. Despite the presence of both consistencies and inconsistencies between the acoustical basis for emotions expression in walking and musical performance, an interpretative is found, which allows to draw a conclusion concerning the validity of the hypothesis by Clynes (1997), according to which emotions expression is based on modality-independent structures: not all of the acoustical parameters considered to test this hypothesis could plausibly be effectively manipulated by the walking-performer. Indeed, while control of average tempo, loudness, and tempo variability was possible and plausibly achievable by each of the participants, this was not the case for articulation and timbre. Concerning articulation, an effective way to achieve control would have been to manipulate the duration of the single footstep sounds, but this was probably hardly possible or simply unachievable. Concerning timbre, the most efficient control could probably have been achieved with a change in the material of the sole. As this opportunity was not given to participants, the timbre of the generated footstep sounds was, more probably, quite uncontrollable, being probably almost completely determined by their weight, and by the material of the soles.

Given these considerations, we are supported in concluding that this study confirmed the modality-independence hypothesis by Clynes (1997), with respect to the acoustical parameters effectively manipulated by walkers: tempo and sound level.

4 Listening experiment

4.1 Methods

4.1.1 Stimuli

A set of 35 stimuli was extracted from the database, choosing, across the five repetitions, one stimulus per walker and walking style. For each walker/emotional intention pair a "stereotypical" walking excerpt was defined by the average value of IOI_{mea}, IOI_{std}, Amp_{mea} and Amp_{std} across the repetitions (see Section 2.1.3.1). For each of the repetitions a Euclidean distance was calculated from the stereotypical walking style, on the basis of the above mentioned acoustical descriptors. The walking excerpt closer to the stereotype was chosen for the experiment. Tables 6-7 report the acoustical properties of the experimental stimuli.

Table 6: Acoustical descriptors extracted from each signal. See text for an explanation of the meaning of each acoustical descriptor.

Walker	Emotion	$T.amp_{mea}$ (dB_{thr})	$H.amp_{mea}$ (dB_{thr})	$T2T_{mea}$ (ms)	$H2T_{mea}$ (ms)	$T.SCG_{mea}$ (Hz)	$H.SCG_{mea}$ (Hz)	$T.mod_{mea}$ (Hz)	$H.mod_{mea}$ (Hz)	$T2H_{mea}$ (ms)	$T.amp_{std}$ (dB_{thr})	$H.amp_{std}$ (dB_{thr})
1	A	17.567	13.16	484.292	48.767	2355.581	3220.998	188.416	161.499	93.913	5.665	7.228
1	F	8.859	11.829	529.475	47.253	3388.943	3106.299	230.713	246.094	321.977	6.547	6.185
1	H	13.968	15.465	487.738	43.774	2892.514	2741.268	275.146	301.465	212.744	7.689	7.819
1	N	8.771	9.335	510.884	90.804	3504.385	3236.052	220.117	165.088	278.923	5.569	4.327
1	S	12.997	9.362	759.574	48.87	2606.512	3401.949	168.677	136.377	419.483	6.428	6.501
2	A	26.042	21.922	437.268	27.175	1978.124	2325.779	430.664	470.654	44.656	2.972	7.679
2	F	21.694	18.917	435.159	16.051	2353.432	2389.186	470.654	372.217	134.305	2.439	8.158
2	H	23.314	17.296	447.54	64.289	1952.579	2728.704	442.969	307.617	105.166	6.853	10.229
2	N	21.831	19.614	474.078	12.805	2429.541	2839.717	446.045	390.674	226.247	3.244	7.463
2	S	18.476	18.82	607.169	10.852	2157.979	2555.582	538.33	538.33	390.125	4.211	4.051
3	A	17.408	18.914	516.489	10.146	2737.636	2527.288	243.018	239.941	313.515	1.745	3.425
3	F	7.571	9.031	991.27	20.385	4104.983	3808.633	304.541	304.541	905.132	1.93	2.884
3	H	19.145	20.771	514.641	10.671	2823.626	2511.757	264.551	261.475	283.844	3.618	3.42
3	N	19.012	19.012	546.209	0	2595.037	2595.037	261.475	261.475	342.475	3.233	3.233
3	S	11.109	13.907	640.729	36.356	3412.896	3067.799	212.256	255.322	473.069	3.369	3.746
4	A	14.167	9.658	489.94	63.67	3211.794	4391.904	313.77	215.332	234.153	3.753	2.583
4	F	9.55	7.722	586.946	62.685	4092.469	5236.231	243.018	243.018	373.205	3.146	2.686
4	H	12.814	8.26	487.109	83.239	3838.403	5846.639	283.008	203.027	251.727	3.507	3.434
4	N	12.917	10.967	538.462	88.419	3830.505	4941.902	239.941	178.418	311.447	2.855	3.414
4	S	9.142	8.849	680.045	18.892	4624.892	4716.557	209.18	215.332	536.047	4.916	5.083
5	A	18.222	19.292	510.548	12.229	2997.472	2776.931	246.094	264.551	238.069	4.712	4.243
5	F	11.702	11.702	810.737	0	3583.429	3583.429	236.865	236.865	670.911	2.317	2.317
5	H	15.264	17.471	579.664	27.804	3389.551	3079.437	276.855	313.77	321.478	6.238	4.248
5	N	17.204	15.902	576.512	38.04	3532.81	3822.657	227.637	209.18	263.277	7.102	7.365
5	S	16.695	16.695	766.107	0	3015.364	3015.364	249.17	249.17	560.775	1.444	1.444
6	A	19.362	19.362	477.585	0	4331.995	4331.995	150.732	150.732	180.506	4.725	4.725
6	F	11.591	10.479	434.138	31.675	5017.011	5686.407	147.656	2030.273	228.137	3.821	5.289
6	H	15.835	15.835	581.145	0	4728.717	4728.717	178.418	178.418	337.162	4.145	4.145
6	N	11.208	11.686	584.18	25.896	5404.493	5978.72	153.809	1135.107	357.37	3.929	3.115
6	S	6.52	7.125	675.839	21.169	5043.777	5036.189	129.199	129.199	547.011	2.993	3.682
7	A	14.51	15.105	510.356	15.694	4107.962	3916.408	481.805	503.339	292.945	4.904	4.119
7	F	5.327	5.327	3159.966	0	4809.23	4809.23	602.93	602.93	3101.083	2.588	2.588
7	H	14.167	14.167	558.186	0	3801.129	3801.129	765.967	765.967	391.304	3.914	3.914
7	N	10.582	10.603	581.289	24.522	4452.981	5241.231	473.73	2713.184	425.779	2.906	2.868
7	S	9.307	9.307	1278.912	0	4190.506	4190.506	305.054	305.054	1188.109	3.466	3.466

Table 7: Acoustical descriptors extracted from each signal. See text for an explanation of the meaning of each acoustical descriptor.

Walker	Emotion	$T2T_{std}$ (ms)	$H2T_{std}$ (ms)	$T.SCG_{std}$ (Hz)	$H.SCG_{std}$ (Hz)	$T.mod_{std}$ (Hz)	$H.mod_{std}$ (Hz)	$T2H_{std}$ (ms)	$p(h)$	Dur (ms)	Nf
1	A	95.086	54.462	821.005	1035.009	70.719	70.952	127.806	0.5	3390.045	8
1	F	272.079	47.117	721.296	585.667	299.849	292.582	266.643	0.571	3245.76	7
1	H	160.315	53.386	1159.373	1219.994	441.795	430.395	127.815	0.444	3990.998	9
1	N	152.48	56.851	1186.735	928.758	231.364	57.98	113.011	0.778	4201.474	9
1	S	254.604	39.063	881.287	1250.615	53.473	57.242	350.745	0.667	3867.574	6
2	A	13.654	46.819	606.372	1209.68	245.201	250.942	39.238	0.286	2623.605	7
2	F	20.653	42.467	779.843	833.455	183.319	177.443	76.149	0.143	2610.952	7
2	H	85.799	63.183	571.332	903.278	214.201	215.588	94.848	0.571	2685.238	7
2	N	12.456	33.88	1070.551	1603.731	118.409	144.143	60.954	0.143	2844.467	7
2	S	11.565	28.712	775.479	1520.327	242.667	242.667	71.787	0.143	3718.98	7
3	A	45.512	26.843	457.68	483.421	46.04	47.224	28.787	0.143	3098.934	7
3	F	227.332	53.935	737.068	621.447	84.711	84.711	218.198	0.143	5947.619	7
3	H	45.931	28.232	660.314	549.797	120.351	121.447	26.721	0.143	3087.846	7
3	N	9.048	0	494.613	494.613	87.405	87.405	32.339	0	3277.256	7
3	S	107.281	62.09	508.087	319.132	72.947	90.019	36.027	0.286	3844.376	7
4	A	30.239	31.821	437.943	789.751	127.911	66.949	85.978	0.857	3028.776	7
4	F	39.673	59.593	436.933	992.695	108.89	105.281	92.436	0.571	3521.678	7
4	H	14.59	39.162	390.97	1141.864	77.068	63.218	24.072	0.857	2922.653	7
4	N	15.757	41.164	222.027	688.247	31.521	56.583	53.367	0.857	3323.764	7
4	S	224.571	49.984	935.277	873.975	29.719	35.164	195.684	0.143	4080.272	7
5	A	57.313	32.354	739.675	350.084	121.72	125.998	109.788	0.143	3063.288	7
5	F	20.763	0	529.374	529.374	113.943	113.943	56.056	0	4864.422	7
5	H	61.284	48.596	668.833	312.894	112.185	136.673	81.25	0.286	3477.982	7
5	N	76.421	65.171	1040.675	1294.23	99.347	135.456	155.052	0.286	3459.07	7
5	S	22.71	0	459.236	459.236	113.165	113.165	31.766	0	4596.644	7
6	A	11.989	0	1104.266	1104.266	48.15	48.15	84.08	0	2865.51	7
6	F	12.664	39.67	988.722	1116.928	60.176	3259.673	25.228	0.429	2604.83	7
6	H	19.432	0	1120.975	1120.975	70.012	70.012	80.978	0	3486.871	7
6	N	14.367	44.931	979.635	976.181	63.912	2605.183	69.11	0.286	3505.079	7
6	S	33.92	56.009	760.287	749.699	51.259	51.259	89.547	0.143	4055.034	7
7	A	70.992	44.391	1216.356	854.622	453.87	436.157	57.872	0.125	3572.494	8
7	F	2110.788	0	1660.434	1660.434	413.359	413.359	2124.8	0	12639.864	5
7	H	31.085	0	539.532	539.532	370.858	370.858	43.436	0	3349.116	7
7	N	8.511	42.008	921.162	2128.43	425.61	5388.81	90.105	0.286	3487.732	7
7	S	558.656	0	783.203	783.203	46.016	46.016	556.267	0	6394.558	6

Table 8 shows the correlation between walker gender, walker weight, shoe length and sole hardness within the experimental set. These were initially assumed to be taken into account by participants when judging walker gender, weight, shoe largeness and sole hardness, respectively. Walker gender was coded as a binary variable (female = 0; male = 1), while sole hardness was conceived as an ordinal variable. The association between continuous variables was tested using the Pearson's product moment correlation, that between binary and continuous variables was assessed using the point biserial correlation, that between ordinal and continuous variables was assessed using Spearman's ρ rank correlation.

Table 8: Correlations among walker parameters within the experimental set. GEN = walker gender; WEI = walker weight; LEN = shoe length; HAR = sole hardness. Significant correlations ($p< 0.5$; df=33) are shown in bold face.

	GEN	WEI	LEN
WEI	**0.691**		
LEN	**0.677**	0.326	
HAR	**-0.663**	**-0.430**	**-0.854**

4.1.2 Procedure

Participants were asked to judge walking sounds with respect to a subset of the measured source properties. A technique similar to the semantic differential (Osgood, Suci, & Tannenbaum, 1957) was used, according to which stimuli are rated along a set of bi-polar scales, defined by one adjective and its contrary (e.g., hard – soft). In particular, the VAME variant was used (Hajda, Kendall, Carterette, & Harshberger, 1997), according to which stimuli were judged along bi-polar scales defined by one adjective and its negation (e.g., hard – not hard). A first group of scales comprised five rating scales related to the mood of the walker. Participants were thus asked to judge how sure they were that the mood of the walker was angry, fearful, happy, normal and sad, where the extremes of each of the scales was defined by sentences of the form "The mood of the walker is XX..." – "... is not XX". A second group comprised four scales referred to the walker, participants being asked to judge how sure they were that the walker was female, male, light and heavy. A third group comprised four scales referred to the shoes, participants being asked to judge how sure they were that the shoes were large and small, and how sure they were that the soles were hard and soft. Ratings were given moving horizontally one slider for each of the scales.

For each of the walking sounds the thirteen sliders were presented together on the screen, in random order from top to bottom. The initial position of the slider was chosen randomly. Participants played the stimulus clicking on a button on the screen and were encouraged to replay it several times while rating it along the different scales. Stimuli were presented in random order, and were judged once along each of the scales.

Before the experiment participants were asked to listen to each of the 35 walking sounds at least once in order to familiarize with the acoustical variation they would have met during the experiment. The entire procedure was practiced with a smaller set of four stimuli, not included in the main experiment.

Stimuli where presented through AKG K240 headphones, connected to the output of the Sound Blaster Live soundcard of the PC used to program the experiment. Participants sat inside an acoustically isolated room. The peak levels of the signals ranged from 32.7 to 62.2 dB SPL.

4.1.3 Participants

Thirteen listeners took part in the experiment on a voluntary basis (age: 25-59; 7 females, 7 males). For each of the listeners the average number of hours of every-day musical listening, the number of years of musical practice, the main played musical instrument and the actual status as musician (practicing vs. not practicing) were ascertained.

4.2 Results

4.2.1 Scales correlation

First, correlations between ratings on the different scales were analyzed. In particular, for each participant the association between the male/female, heavy/light, large/small and hard/soft ratings was ascertained, using the Pearson's product moment correlation r, with the purpose of assessing the plausibility of combining data from the contrasted judgment scales.

For all participants but one significative negative correlations between all the compared rating scales were observed (max $r = -0.527$; average $r = -0.842$; $\mathrm{p} \leq 0.001$; $\mathrm{df} = 33$). Data for one of the participants was characterized, in general, by weaker correlations (male/female $r = 0.148$; heavy/light $r = -0.311$; large/small $r = -0.426$; hard/soft $r = -0.586$; $0.397 \geq \mathrm{p} \leq 0.001$; $\mathrm{df} = 33$). Data from this participant were not further considered. For the remaining participants ratings given for the opposed scales were combined. Thus, a maleness scale was defined as $(m_i - f_i + 1)/2$ where m_i and f_i are the rating given on the male and female scales for the i^{th} stimulus, respectively. Similarly, the walker heaviness, shoes largeness and shoes sole hardness scales were derived. Table 9 shows the correlation among rating scales.

As can be seen the strongest correlations are found among mood-related and among walker-related scales. In particular, higher happy ratings are associated with lower fearful and sad ratings and with higher normal ratings; higher fearful ratings are associated with lower normal ratings; higher angry ratings are associated with lower angry ratings. All the walker-related scales are found moderately to strongly correlated. In particular, the maleness, walker heaviness and shoe largeness were positively correlated, walkers perceived as male being perceived also as heavier and as wearing larger shoes, and the sole hardness scale was negatively correlated with all the other walker-related scales, where harder soles where perceived as being used by light, female walkers wearing smaller shoes. It should be noted that this latter group of correlations maps well the correlation among actual walker properties (see Table 8). Correlations between mood- and walker-related scales where, instead, moderated at best. In particular, perceived male walkers were also perceived as less angry and sadder, while walking sounds perceived as generated with harder soles were perceived as generated by angrier walkers. Figure 9 shows, for each of the mood-related

Table 9: Correlations among rating scales averaged across participants. ANG = anger; FEA = fear; HAP = happiness; NOR = "normalness"; SAD = sadness; MAL = maleness; HEA = walker heaviness; LAR = shoe largeness; HAR = sole hardness. Significant correlations ($p< 0.5$; df=33) are shown in bold face.

	ANG	FEA	HAP	NOR	SAD	MAL	HEA	LAR
FEA	0.010							
HAP	0.233	**-0.491**						
NOR	-0.109	**-0.826**	**0.715**					
SAD	**-0.770**	0.223	**-0.680**	-0.321				
MAL	**-0.459**	-0.103	-0.116	0.107	**0.393**			
HEA	-0.207	-0.224	-0.104	0.132	0.159	**0.876**		
LAR	-0.320	-0.250	-0.080	0.188	0.250	**0.931**	**0.960**	
HAR	**0.339**	-0.065	-0.032	-0.018	-0.226	**-0.784**	**-0.577**	**-0.633**

rating scales, the rating averaged across participants for each of the walking sounds as well as the average rating for each of the actual emotional intentions.

Comparison of the results from the different mood-related scales show that, in general, the lower ratings were observed for the happiness scale, indicating that the investigated sounds were particularly ineffective in mediating perception of a happy mood. Inspection of data from each of the mood-related scales reveal that for the anger, fear and sadness scales the highest ratings were given, on the average, to the angry, fearful and sad walking sequences, respectively, while for the happy and normal scales the highest scores were observed for the angry and happy scales, respectively.

Figure 10 shows, for each of the walker-related scales, the rating averaged across participants for each of the walking sounds as well as the average rating observed for each of the level of the related walker property.

As can be seen, exception done for the walker heaviness one, for each of the walker-related scales average ratings map closely actual walker properties. For the walker heaviness scale, instead, ratings are an inverted U-shaped function of actual walker weight. It should furthermore be pointed out that this result is not an artefact of averaging results from participants with markedly different response profiles (this result would indeed be observed if for half of the participants heaviness decreased with increasing walker weight and increased for the other half), but characterizes ratings of the vast majority of participants. Given this result, and the singularity of the observed psychophysical function for sound source recognition research (Giordano, 2005), it might be hypothesized that participants estimated walker heaviness using focusing, mistakenly, on some other source property. Accordingly, it was tested which of the measured walker properties reported in Table 1 had the strongest association with perceptual weight estimates. Correlation coefficients were used to this purpose. The same test was also carried for the other walker-related rating scales. In

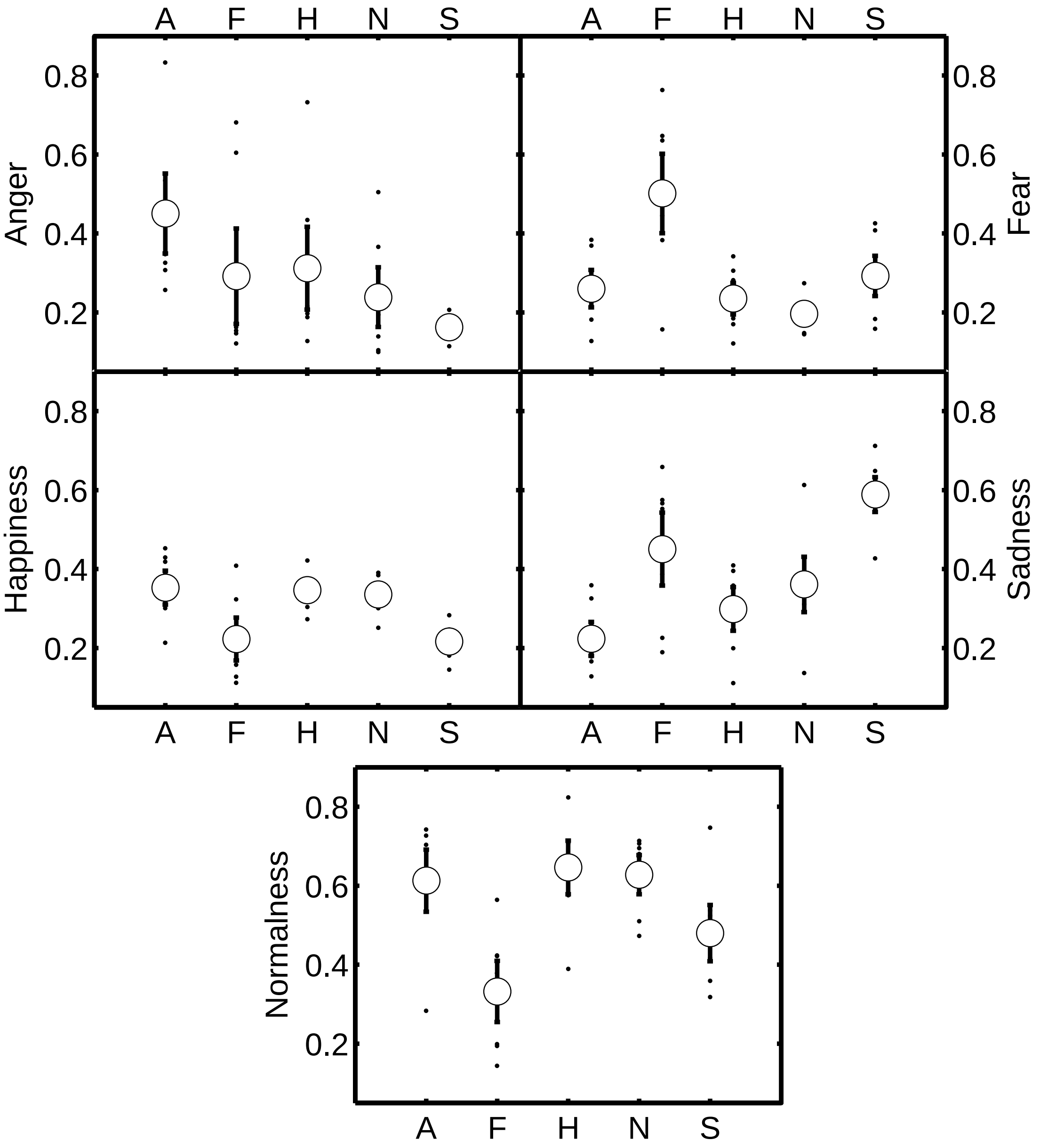

Figure 9: Mood ratings as a function of the actual emotional intention for each of the walking sounds (black dots). A = angry; F = fearful; H = happy; N = normal; S = sad. White circles show the average rating for each of the walking intentions. Error-bars = ± 1 SE.

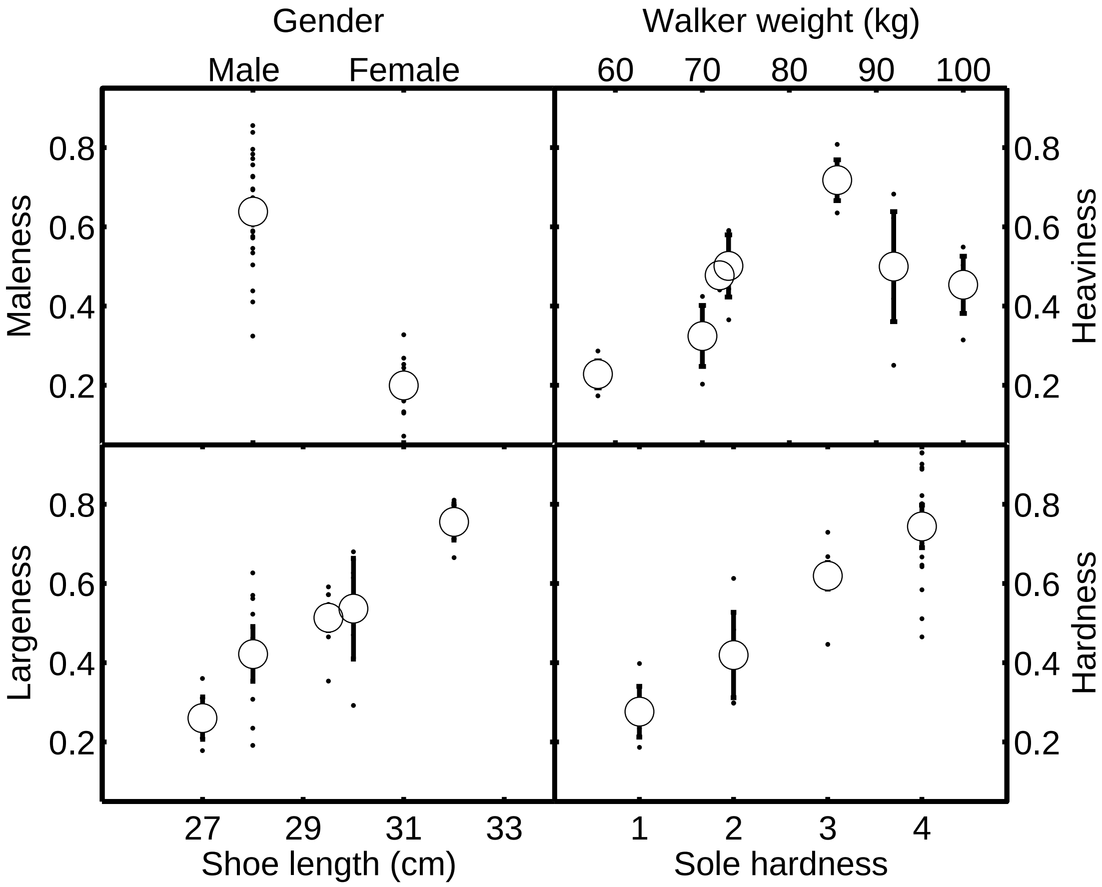

Figure 10: Walker-related ratings as a function of the actual walker properties for each of the walking sounds (black dots). White circles show the average rating for each of the levels of the considered walker descriptor. Error-bars = ± 1 SE.

any case ratings averaged across participants were considered. When the association between a continuous walker variable and the rating was assessed, the Pearson's product moment correlation, which tests for the presence of a linear relationship between variables, was used. The association of ratings with gender was assessed using the point biserial correlation, gender being coded as a binary variable (female = 0; male = 1). The association of ratings with sole hardness was assessed using Spearman's ρ rank correlation, where sole hardness was coded as an ordinal variable[1]. Table 10 reports the results of this analysis.

---

[1] Spearman's rank correlation should test for a monotone relationship between the two variables

Table 10: Correlations between walker properties and average ratings along each of the walker-related scales.

	Rating scale			
Walk. Prop	Maleness	Heaviness	Largeness	Hardness
Age	-0.513	-0.326	-0.373	0.498
Gender	**0.860**	0.691	0.744	-0.720
Weight	0.653	0.500	0.578	-0.638
Height	0.393	0.330	0.415	-0.008
Ground-Hip	-0.023	0.026	0.071	0.374
Ankle-Knee	0.190	0.372	0.352	0.088
Knee-Hip	-0.413	-0.394	-0.432	0.360
Shoe Length	0.773	**0.818**	0.788	-0.654
Shoe Width	0.799	0.790	**0.791**	-0.653
Sole Hardness	-0.689	-0.643	-0.613	**0.766**

As can be noted, the walker variable most strongly associated with the maleness rating scale is gender, that most strongly associated with perceived walker weight is shoe length, that most strongly associated with perceived shoe largeness is shoe width, and that most strongly associated with rated sole hardness is sole hardness. Thus, it might be concluded that, in order to estimate walker weight, participants focused on the properties of the shoes, most probably shoe length.

4.2.2 Recognition performance

For each participant recognition performance was evaluated using a measure, RP, common to the different rating scales.

For each rating scale ψ, associated to a given stimulus measure ϕ, the observed probability $p_{(i,j)}$ was calculated, given by the proportion of times that stimuli for which $\phi = i$ received a higher rating than stimuli for which $\phi = j$. It should be noted that, under the assumption that all the ratings given for the stimuli for which $\phi = i$ differ $p_{(i,i)} = 0.5$, that $p_{(i,j)} = 0.5$ if the levels/categories i and j are not discriminated by the judgment and that $p_{(i,j)} + p_{(j,i)} = 1$.

With the maleness scale good performance is observed if $p_{(M,F)} > 0.5$ or, equivalently, if $p_{(F,M)} < 0.5$ and that consistent misidentification (i.e., all female walking sounds are always rated as more male than the male walking sounds) is observed if $p_{(M,F)} < 0.5$. In this case $RP = p_{(M,F)}$.

For the walker weight, shoe size and sole hardness scales the considered stimuli properties were walker weight, shoes length and sole hardness, respectively. In any case good performance is found if $p_{(i,j)} > 0.5$ with $i > j$ or, equivalently, if $p_{i,j} < 0.5$ with $i < j$. Misidentification is instead observed if $p_{(i,j)} < 0.5$ with $i > j$. RP is thus taken to be equal to the average of all the proportions $p_{(i,j)}$ for which $i < j$, i.e., $RP = \overline{p_{(i,j)}}$ with $i < j$.

Concerning emotional intention recognition, given a rating scale ψ and the associated emotional walking style *emo* (e.g., $\psi =$ sadness scale, $emo =$ sad walking style), good performance is found if $p_{(emo,j)} > 0.5$ or equivalently if $p_{(j,emo)} < 0.5$, with $j \neq emo$. Thus, for these scales $RP = \overline{p_{(emo,j)}}$ with $j \neq emo$. An overall measure of emotional intention recognition was defined as the average of the RP measures computed for each of the emotional rating scales. A final observation should be made concerning the RP measure defined for the mood-related rating scales. It should be noted that in this case a good recognition performance should not only assign higher ratings to the *emo* walking styles, but should strongly confuse all the other walking styles with each other, i.e., $p_{(i,j)} = 0.5$ with $i \neq j$ and $\{i, j\} \neq emo$. Ideally, also these latter proportions should be included in the definition of RP. However, if simply averaged with $p_{(emo,j)}$ they would force RP close to 0.5, the measure assumed to characterize perfect confusion. For this reason and in order to avoid complications in the definition of RP, this latter group of proportions was not considered.

To summarize, with all the rating scales a measure of recognition performance was defined, RP, where $RP > 0.5$ characterizes perfect performance, $RP = 0.5$ characterizes perfect confusion among the stimuli levels with respect to the rating scale, and $RP < 0.5$ characterizes consistent misidentification.

For each scale the hypotheses $RP = 0.5$, $RP > 0.5$ and $RP < 0.5$ where tested for the entire group of participants, i.e., for each scale it was tested whether RP averaged across participants differed significantly from 0.5 and if it was higher or lower than this value. The bootstrap approach was used to this purpose (Efron & Tibishirani, 1993). BC_a 95% bootstrap confidence intervals for the mean RP averaged across participants were computed. Bootstrap confidence intervals were computed on the basis of 5000 bootstrap samples, drawn resampling participants.

Table 11 shows, for each judgment scale, the value of RP averaged across participants. Also, the limits of the bootstrap BC_a 95% confidence interval for the RP mean is shown. The same data is shown in Figure 11. As can be noted, all recognition performance measures are significantly higher than 0.5, i.e. recognition of all investigated attributes of the walking event was better than chance.

An interesting issue then concerns whether any of the collected measures of participants' musical expertise (hours of daily musical listening, years of musical practice, current practice with the instrument, played instrument), or participants age and gender, explains performance with the different judgment scales. Given the informal nature of this assessment, statistical modeling was limited to simple one-way ANOVA models, where each of the participants descriptors were used, in turn, as independent, and the RP derived for each of the rating scales were used as dependent. Also used as dependent were the RP for overall emotional intention recognition, and an overall recognition performance measure, computed averaging RP for each participant across all the rating scales. Continuous participants' descriptors were transformed into binary variables, dissecting the distribution of values in two portions containing, approximately, an equal amount of participants (6 and 7 participants). Thus tested, few participant de-

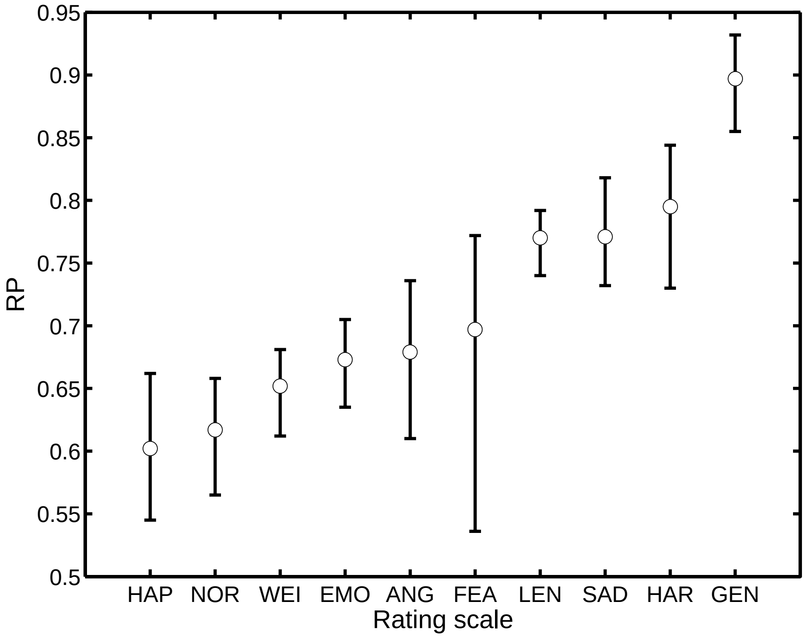

Figure 11: Analysis of recognition performance for all the judgment scales. Error bars bracket 95% BC_a bootstrap confidence intervals for the RP average.

Table 11: Analysis of recognition performance. Average RP values averaged across participants and bootstrap $BC_a 95\%$ confidence intervals (c.i.) for the mean. Scores significantly higher than 0.5 are indicative of better than chance recognition performance.

Scale	RP	95% c.i.
Anger	0.679	0.610/0.736
Fear	0.697	0.536/0.772
Happiness	0.602	0.545/0.662
Normal mood	0.617	0.565/0.658
Sadness	0.771	0.732/0.818
Overall emotions	0.673	0.635/0.705
Gender	0.897	0.855/0.932
Weight	0.652	0.612/0.681
Length	0.770	0.740/0.792
Hardness	0.795	0.730/0.844

scriptors were found associated with differences in recognition performance for a limited number of rating scales. In particular, females were found having a significantly better recognition performance for all the rating scales but the sole hardness one. However, this difference was significant only for the anger rating scale and for recognition of emotion in general ($F(1, 11) \geq 5.959$, p$\leq$ 0.033). Younger participants (age$\leq$27) were found having significantly better performance in recognition of anger ($F(1, 11) = 12.106$, p= 0.005). Participants that daily listened to music for a shorter time (1 hour or less) were found better at recognizing sole hardness ($F(1, 11) \geq 5.212$, p$\leq$ 0.043). Finally, participants that currently practiced a musical instrument were found having superior recognition performance with all the scales but with the anger- and gender-related ones. However, this difference was significant only for the scale which tested recognition of the emotionless walking style ($F(1, 11) \geq 5.959$, p$\leq$ 0.033).

4.2.3 Acoustical criteria for source perception

The acoustical criteria used by participants to evaluate the different properties of the walking sound source were investigated in a regression framework.

It was assumed that source properties were evaluated using one or more acoustical features at the same time, and that the transform relating levels of the acoustical features to the estimated source properties was monotone. Thus, for each of the judgment scales, the first step has been to highlight such a monotone transform, limiting the possibilities to functions of the form $\psi = a + b\alpha^c$, where ψ is the behavioral response, and α is the acoustical feature. The parameters a, b and c were estimated by means of an iterative least squares minimization procedure. Following the advices by Hosmer and Lemeshow (1989), before building multivariate regression models for the behavioral variables, the significance of the association between the transformed predictor $a+b\alpha^c$ and the behavioral variable ψ was tested. Transformed predictors were thus considered in a multivariate framework only if the univariate model was found significant. Finally, a tradeoff between goodness-of-fit of the acoustically-based model to the data, and the simplicity of the model was sought. Thus, the final models were considered to correspond to the most economical models, i.e. with the fewer number of predictors, whose goodness-of-fit was above a given threshold, i.e. whose adjusted R^2 (R^2_{adj}) was equal or higher than 0.7 and thus explained more than 70% of the variance in observed data. If this procedure produced more than one model, i.e. highlighted more than one above-threshold model with the same/minimal number of predictors, these were considered as statistically equivalent and as representing equally plausible alternative acoustical explanations for the modeled behavioral variable. If for a given behavioral variable no model was found with a higher-than-threshold R^2_{adj}, then all the significant univariate models were retained. This outcome had to be taken as representing a failure in giving a satisfactory acoustical explanation for perception. Nonetheless, those models associated with the higher R^2_{adj} values gave indications concerning which of the extracted acoustical features were, in part, at the basis of experimental judgment.

None of the univariate models exceeded the R^2_{adj} threshold. Table 12 reports the statistics for those univariate which included a significant effect of the considered acoustical descriptor.

Table 12: Acoustical descriptors found significantly associated with the source-related judgment scales, as tested with univariate regression models. For each descriptor/scale pair the R^2_{adj} goodness-of-fit measure is reported in italic, along with the exponent c estimated with a least squares minimization procedure. The sign of the relationship between the acoustical descriptor and the judged source attribute is also shown, a "+" that increasing levels of the acoustical descriptor are associated with increasing estimates of the source property.

Angry	$T.amp_{mea}$	$T2H_{mea}$	Dur	$H.amp_{mea}$	$T2T_{mea}$	$H.amp_{std}$	$T.SCG_{mea}$	$H.SCG_{mea}$	
	3.414	*2.145*	*-0.968*	*-0.542*	*-1.124*	*-1.043*	*6.894*	*-1.262*	
	0.695	0.659	0.502	0.411	0.41	0.393	0.336	0.164	
	+	-	-	+	-	-	-	-	
Fearful	$T2H_{std}$	$T.amp_{mea}$	Nf	$T2T_{mea}$	Dur	$T2T_{std}$	$H.amp_{mea}$		
	1.693	*0.516*	*-2.14*	*-2.578*	*-0.001*	*0.719*	*-2.108*		
	0.245	0.232	0.228	0.196	0.181	0.174	0.158		
	-	-	-	+	+	+	-		
Happy	$T2T_{mea}$	Dur	$T2H_{std}$	$T2T_{std}$	$T.amp_{mea}$	$T2H_{mea}$	$H.amp_{mea}$	Nf	
	-0.001	*0.003*	*-0.001*	*0.002*	*-0.875*	*-0.001*	*-0.893*	*0.007*	
	0.465	0.431	0.321	0.309	0.223	0.202	0.142	0.116	
	-	-	-	-	+	+	+	+	
Normal	$T2H_{std}$	$T2T_{std}$	$H.amp_{mea}$	$T2T_{mea}$	$T.amp_{mea}$	Nf	Dur		
	0.002	*-0.003*	*-0.001*	*0.003*	*-0.001*	*-0.003*	*0.008*		
	0.238	0.213	0.185	0.174	0.167	0.141	0.128		
	+	+	+	-	+	+	+		
Sad	$T.amp_{mea}$	$T2H_{mea}$	Dur	$H.amp_{mea}$	$T2T_{mea}$	$T2T_{std}$	$T2H_{std}$	$H.amp_{std}$	$T.SCG_{mea}$
	-0.004	*0.002*	*0.004*	*0.002*	*0.003*	*0.002*	*-0.005*	*0.05*	*-0.004*
	0.528	0.517	0.465	0.407	0.369	0.208	0.19	0.186	0.158
	-	+	+	-	+	+	+	+	+
Male	$T.mod_{mea}$	$H.mod_{mea}$	$T.mod_{std}$	$H.mod_{std}$					
	-0.004	*-1.231*	*-0.003*	*-0.697*					
	0.597	0.372	0.122	0.093					
	+	+	+	+					
Large	$T.mod_{mea}$	$H.mod_{mea}$	$T.amp_{std}$	Nf	$p(h)$	$H.mod_{std}$			
	-0.003	*-0.949*	*1.413*	*-0.369*	*0.927*	*4.076*			
	0.419	0.375	0.244	0.194	0.096	0.092			
	+	+	+	+	+	+			
Heavy	$T.mod_{mea}$	$H.mod_{mea}$	$T.amp_{std}$	Nf					
	-0.003	*-0.87*	*-0.002*	*8.193*					
	0.363	0.356	0.174	0.17					
	+	+	+	+					
Hard	$T.mod_{mea}$	$T.SCG_{mea}$	$H.SCG_{mea}$	$T.amp_{mea}$	$T.mod_{std}$				
	-0.002	*-0.82*	*-0.774*	*0.003*	*0.002*				
	0.623	0.308	0.19	0.146	0.101				
	+	+	+	+	-				

Several above-threshold models were found. Each of them included two predictors and modeled either the anger or sole hardness judgments. Table 13 summarizes the final above-threshold models. These models take the form $\psi = a + b_1\alpha_1^{c_1} + alpha_2^{c_2}$. The parameters $c_{1,2}$ are reported in a standardized form, where the higher the absolute value of the standardized parameter estimate the higher the relevance of the associated acoustical descriptor in determining the modeled behavioral response.

Table 13: Final models selected to explain perceptual estimation of the walking sounds source properties. For all models it is shown the modeled judgment, the acoustical attributes in the model ($\alpha_{1,2}$), the standardized parameter estimates for the effect of the acoustical attributes ($b_{1,2}$) and the R^2_{adj} goodness-of-fit measure.

Attribute	α_1	b_1	α_2	b_2	R^2_{adj}
Angry	$T.amp_{mea}$	0.638	Dur	0.330	0.759
Angry	$T.amp_{mea}$	0.513	$T2H_{mea}$	0.401	0.744
Angry	$T.amp_{mea}$	0.687	$T2T_{mea}$	0.278	0.743
Angry	$T.amp_{mea}$	0.995	$H.SCG_{mea}$	-0.232	0.717
Angry	$T.amp_{mea}$	0.725	$H.amp_{std}$	0.179	0.706
Angry	$T.amp_{mea}$	1.030	$H.amp_{mea}$	-0.223	0.700
Angry	$T.amp_{mea}$	0.991	$T.SCG_{mea}$	-0.191	0.700
Hard	$T.mod_{mea}$	0.679	$T.SCG_{mea}$	0.337	0.718
Hard	$T.mod_{mea}$	1.101	$T.mod_{std}$	-0.428	0.707

Figure 12 shows the best multivariate models found accounting for judgments of walker anger and of shoe sole hardness.

4.3 Discussion

A listening test was carried, were participants were asked to evaluate several different attributes of the walkers, emotion included. Two main groups of strongly correlated judgments were found: those related to the emotion on the one hand and those concerning all the other walkers properties on the other. Thus, emotion perception in walking was found almost orthogonal with respect to perception of walker attributes gender, weight and shoe size and shoe sole hardness. Particularly strong correlations were found among these latter four perceptual continua, correlations that mapped the actual association among these parameters in the experimental stimulus set. Care should be taken, in a following study, to test whether correlated perceptions are caused by correlations of source properties in the stimulus set, or are instead given by a high order cognitive parameter which takes advantage to regularities in the environment (c.f. Barlow, 2001).

A statistical test was developed to compare performance in recognition of the different walker attributes. Results highlighted above-chance performance for all the properties, both emotion-related and not. Walker emotion, in general, was among the worst recognized attributes, even though performance varied strongly across the different emotions sadness being the better recognized, happiness the worst. Non-emotion related parameters were, instead, the better recognized, where the best performance was found for gender and for shoe sole hardness. Consequently, following the arguments advanced in Section 1, it might

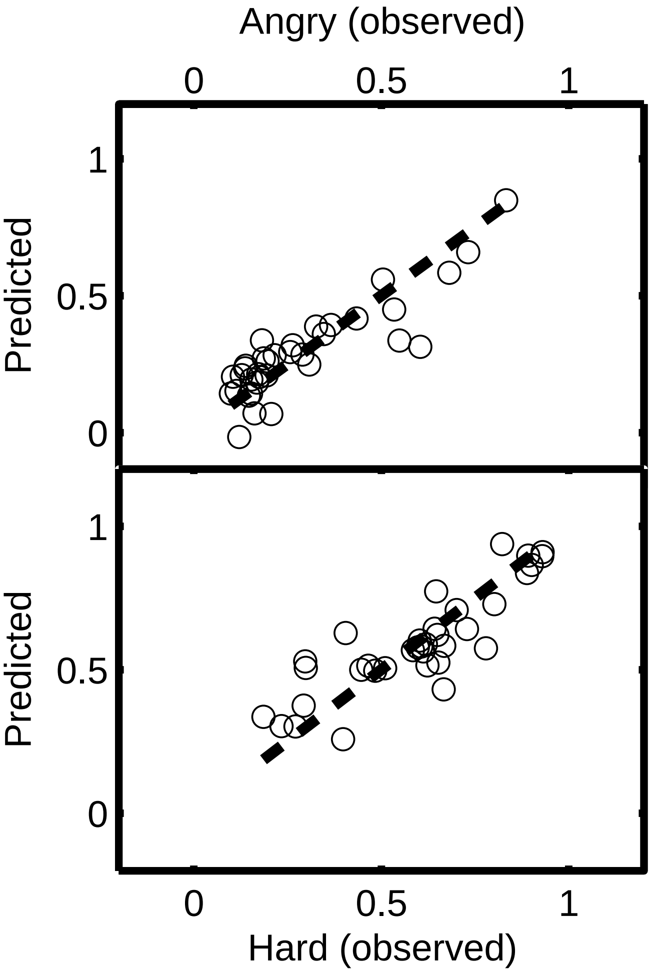

Figure 12: Best regression models found explaining judgment of walker anger (top panel) and of shoe sole hardness (bottom panel). Judgments predicted by the regression models is shown as a function of the observed judgment. The dashed line highlights the condition of perfect correspondence of observed and predicted data.

be concluded that these latter two attributes possess the highest relevance to the everyday walking sounds listener.

Several listeners measures were tested in their power to explain recognition performance. Notably, measures of musical practice virtually explained performance with none of the recognized attributes, while listeners gender was found associated to significant differences in emotions recognition performance, females being better than males.

Finally, acoustical criteria for recognition were outlined. In general emotion recognition was found mainly based on sound level- and tempo-related properties, while recognition of non emotions-related attributes was found based on spectral properties of the walking sounds. However, models which explained significantly well observed data were found for only two of the investigated attributes: anger and shoe sole hardness. For anger the best explanatory model was based on the average level of the walking excerpts and on signal duration, while shoe sole hardness was better explained by the spectral mode and centroid of the footstep sounds.

References

Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral & Brain Sciences*, *24*(4), 602-607.

Bresin, R., & Battel, G. U. (2000). Articulation strategies in expressive piano performance. analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart's sonata in G major (K 545). *Journal of New Music Research*, *29*(3), 211-224.

Bresin, R., & Dahl, S. (2003). Experiments on gestures: walking, running, and hitting. In D. Rocchesso & F. Fontana (Eds.), *The sounding object* (p. xx-xx). Firenze: Mondo Estremo.

Bresin, R., & Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal*, *24*(4), 44-63.

Clynes, M. (1997). *Sentics: The touch of emotions.* New York: Anchor Press/Doubleday.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and Psychological Measurement*, *33*, 107-112.

Efron, B., & Tibishirani, R. J. (1993). *An introduction to the bootstrap.* New York, NY: Chapman & Hall.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68-91.

Giordano, B. L. (2005). *Sound source perception in impact sounds.* Unpublished doctoral dissertation, University of Padova, Italy.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliege & J. Sloboda (Eds.), *The perception and cognition of music* (p. 253-306). London: L. Erlbaum.

Hartmann, W. M. (1997). *Signals, sound and sensation.* Woodbury, NY: AIP Press.

Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Juslin, P. (2001). Communicating emotion in music performance: a review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (p. 305-333). New York: Oxford University Press.

Juslin, P. N. (1997a). Emotional communication in music performance: a functionalist perspective and some data. *Music Perception*, *14*(4), 383-418.

Juslin, P. N. (1997b). Perceived emotional expression in synthesized performances of a short melody: capturing the listener's judgment policy. *MusicæScientiæ*, *1*(2), 225-256.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770-814.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 279-294.

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *J. Acoust. Soc. Am.*, *90*(6), 3036-3049.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: the cognitive psychology of human audition* (p. 146-198). Oxford University Press.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomecanics of simulated sound sources: Material properties of impacted bars. *J. Acoust. Soc. Am.*, *115*(3), 1306-1320.

Nilsson, J., & Thorstensson, A. (1987). Adaptability in frequency and amplitude of leg movements during human locomotion at different speeds. *Acta Physiol Scand.*, *129*(1), 107-114.

Nilsson, J., & Thorstensson, A. (1989). Ground reaction forces at different speeds of human walking and running. *Acta Physiol Scand.*, *136*(2), 217-227.

Osgood, C. E., Suci, G. H., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

Pastore, R. E., Flint, J. D., Gaston, J. R., & Solomon, M. J. (submitted). The source - perception loop for posture in human gait.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 704-712.