Applied Acoustics 75 (2014) 59–66

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Semantic congruence in audio–haptic simulation of footsteps

Luca Turchet *, Stefania Serafin

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

ARTICLE INFO

Article history:
Received 5 December 2012
Received in revised form 17 June 2013
Accepted 24 June 2013
Available online 3 August 2013

Keywords:
Audio
Haptic
Semantic congruence

ABSTRACT

In this paper we present an experiment whose goal is to investigate subjects' ability to match pairs of synthetic auditory and haptic stimuli which simulate the sensation of walking on different surfaces. In three non-interactive conditions the audio–haptic stimuli were passively presented through a desktop system, while in three interactive conditions participants produced the audio–haptic feedback interactively while walking. Results show that material typology (i.e., solid or aggregate) is processed very consistently in both the auditory and haptic modalities. Subjects expressed a higher level of semantic congruence for those audio–haptic pairs of materials which belonged to the same typology. Furthermore, better matching ability was found for the passive case compared to the interactive one, although this may be due to the limits of the technology used for the interactive haptic simulations.

1. Introduction

In the field of multimodal perception and cognition, several studies have investigated the reactions of subjects when presented simultaneously with stimuli in two different modalities. Studies on the audiovisual multisensory integration show that the binding of the two modalities depends on several factors [1]. One of those is whether the stimuli are semantically congruent or not [2,3]. This means that subjects are able to provide a congruent meaning to the stimuli in both modalities. An example of a semantically congruent stimulus in the audio-visual domain is seeing the picture of a dog and hearing a barking sound. The same bimodal stimulus would be semantically incongruent if the picture of a dog was shown while hearing the sound of a cat meowing [2].

Semantically congruent stimuli have been shown to enhance behavioral performance [3,2], and to aid the identification of masked images [4]. However, further investigations are needed to determine under which conditions semantic congruency influences audiovisual multisensory integration [5].

On the other hand, several studies have been conducted on multimodal perception involving the auditory and haptic modalities [6–9]. However, to our knowledge few previous studies investigated the semantic congruence between audition and touch. This is especially the case when auditory and tactile stimuli are presented at feet level since research on the interaction between touch and audition has focused mainly on the hand [8,9].

Although the foot–ground interactions are phenomena which produce rich sensory information, few studies on both the auditory and haptic perception have been conducted in this context. At auditory level, Li and colleagues investigated the ability of subjects to identify the gender of a human walker by listening his/her footstep sounds [10], while Pastore and co-workers investigated listeners' ability to make judgments about the posture of the walker who generated the acoustic stimuli [11].

Moreover, the haptic perceptual system has been proven to be able to discriminate grounds of different elasticity while walking [12], and the vibrotactile sensory channels showed to play an important role in the perception of ground surface compliance during walking [13].

Furthermore, the interaction of auditory and haptic feedback in foot has been studied in [14], showing that the feet were also effective at probing the world with discriminative touch, with and without access to auditory information.

Recently, we developed an interactive system which can provide combined auditory and haptic sensations that arise while walking on solid and aggregate surfaces (the latter being assumed to possess a granular structure, such as that of gravel). The system is composed of an audio–haptic synthesis engine, and a pair of shoes enhanced with sensors and actuators able to provide plantar cutaneous vibration feedback. Such system can be used also non-interactively, providing the user with the audio–haptic feedback while sitting on a chair.

The ecological validity of the auditory as well as of the haptic stimuli involved in the present experiment was assessed in previous research. The results of two non-interactive listening experiments showed that the majority of the simulated surfaces was recognized with high accuracy [15,16]. In particular they were proven to be correctly classified in the corresponding solid and aggregate surface typology. Similar accuracy was noticed in the

* Corresponding author.
E-mail addresses: tur@create.aau.dk (L. Turchet), sts@create.aau.dk (S. Serafin).

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.06.016

recognition of real recorded footstep sounds, which was an indication of the success of the proposed algorithms and their control. Analogously, results of a haptic recognition experiment involving the proposed non-interactive system indicated that subjects were able to correctly discriminate the typology of the simulated surfaces [16]. Moreover, the proposed non-interactive haptic feedback has been recently proven to significantly increase the realism of the desktop-simulated walking experience [17].

Furthermore, the ecological validity was also assessed for the interactive simulations, yielding results similar to those reported for the non-interactive case [15,18,17].

Moreover, from a comparison between the results of the auditory and haptic conditions in [16] and [18], it is possible to notice that participants were able to correctly categorize in both modalities the typology of the simulated surface materials (i.e., solid or aggregate). Indeed solid materials were rarely confused with aggregates, and vice versa, and participants, when not recognizing the presented surface material tended to classify it as another belonging to a same typology (e.g., wood–concrete, snow–frozen snow, dry leaves–forest underbrush) rather than to different typologies (e.g., wood–gravel, metal–dry leaves).

Similar results were found in a recognition task involving the walking on real materials, proving the ability of humans to distinguish almost perfectly between solid and aggregate materials both at auditory and haptic level [14].

Consequently, all these findings suggest that material typology is processed very consistently in the two modalities. In order to investigate the extent to which the two modalities are similar, we conducted an experiment in which we assessed the capacity of subjects to associate the auditory and foot-haptic stimuli provided both in semantically congruent and incongruent way, and both interactively and non-interactively.

To the best of our knowledge, the study of the semantic congruence between the auditory and foot-haptic modalities is a research topic still unexplored. The present research is relevant for the study of the multisensory perception of material properties during walking, for the understanding of the mechanisms underlying the multisensory categorization, and for the comparison of the structure of the perceptual spaces of the audition and foot-touch modalities.

2. Method

We conducted a between-subjects experiment divided in six conditions whose goal was to investigate the ability of subjects to match the different sounds and haptic sensations they were presented with. On three of the six conditions, subjects were not walking but were sitting on a chair and received passively the audio–haptic stimuli (non-interactive conditions). Conversely, in the other three conditions subjects produced the audio–haptic feedback interactively while walking (interactive conditions). The between subjects approach was chosen in order to avoid possible learning effects.

2.1. Apparatus

In previous research we developed a system which simulates both non-interactively and interactively the auditory and haptic sensation of walking on different surfaces [19,20]. To this purpose, shoes enhanced with actuators and pressure sensors were developed. The shoes were a pair of light-weight sandals (Model Arpenaz-50, Decathlon, Villeneuve d'Ascq, France). This particular model was chosen since it has light, stiff foam soles where it is relatively easy to insert sensors and actuators. Four cavities were made in the sole to accommodate four vibrotactile actuators (Haptuator, Tactile Labs Inc., Deux-Montagnes, Qc, Canada). These electromagnetic recoil-type actuators have an operational, linear bandwidth of 50–500 Hz and can provide up to 3 G of acceleration when connected to light loads [21]. In each shoe, two actuators were placed under the heel and the other two under the toe. They were bonded in place to ensure good transmission of the vibrations inside the soles. When activated, vibrations propagated far in the foam. In addition, the sole had two force sensitive resistors sensors intended to pick the foot-floor interaction force in order to drive the auditory and haptic synthesis. The two sensors were placed in correspondence to the heel and toe respectively in each shoe.

The involved hardware allowed the control in real-time of an audio–haptic synthesis engine based on physical models [19]. For the purpose of the experiment, the engine was set in order to synthesize footstep sounds on both solid and aggregate materials, which were simulated using an impact model [22] and a physically informed sonic model (PhiSM) algorithm [23]. In particular, the same models were used to drive both the haptic and the auditory synthesis. They are briefly recalled below.

In the simulation of impact with solids, the contact was modeled by a Hunt–Crossley-type interaction where the force, f, between two bodies, combines hardening elasticity and a dissipation term [24]. Let x represent contact interpenetration and $\alpha > 1$ be a coefficient used to shape the nonlinear hardening, the special model form we used is:

$$f(x,\dot{x}) = -kx^{\alpha} - \lambda x^{\alpha}\dot{x} \quad \text{if } x > 0, \quad 0 \text{ otherwise}. \tag{1}$$

The model described was discretized as proposed in [22].

To simulate aggregate surfaces, the PhiSM algorithm was adopted. This algorithm simulates particle interactions by using a stochastic parameterization thereby avoiding to model each of many particles explicitly. Instead, the particles are assigned a probability to create an acoustic waveform. In the case of many particles, the interaction can be represented using a simple Poisson distribution, where the sound probability is constant at each time step, giving rise to an exponential probability weighting time between events.

In the interaction between a foot and a sole an important element is the ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. In our simulations the physical models were driven by a signal expressing the GRF, which corresponded to the amplitude envelope extracted from an audio signal containing a footstep sound.

The synthesis engine can work both interactively and non-interactively. In this study to control the engine interactively, we used recorded GRF files corresponding to heel and toe strikes which were triggered according to the shoe sensor activated during the user locomotion [20]. To control the engine non-interactively, we created different audio files using the recording of a single real footstep sound on concrete. The envelope profiles of each step in the file were extracted and fed to the engine which produced the synthesized footstep sounds according to the choice of the surface to be simulated. In particular, the surface profile chosen for the experiments was the flat one. It was simulated by placing the footstep sound generator at equal intervals of time, precisely 750 ms, which corresponds to a moderately fast gait [25].

2.1.1. Setup

The experiment was carried out in an acoustically isolated room where the setup was installed. It consisted of a simple graphical user interface with which participants were asked to interact, a spreadsheet to collect their answers, a set of headphones (Sennheiser HD 600) and the haptic shoes previously described. The interface comprised numbered buttons. A button press triggered the presentation of the audio–haptic stimulus through the headphones and the haptic shoes. Users were asked to press each button according to the numerical order, and to type the corresponding answers on the spreadsheet.

During the interactive conditions, participants walked on a carpet-covered floor.

2.2. Stimuli

The synthesis engine simulated six different audio–haptic walked-upon materials: two solid, (wood and metal), and four aggregate (snow, gravel, sand and dry leaves). Therefore, four typologies of pairs occurred by combining the materials at auditory and haptic level:

- solid–solid (e.g., wood–metal)
- aggregate–aggregate (e.g., sand–snow)
- solid–aggregate (e.g., wood–dry leaves)
- aggregate–solid (e.g., gravel–metal)

The audio–haptic displays selected for this study simulated materials of different degrees of compliance (from the least to the most compliant: metal, wood, dry leaves, sand, gravel, snow) that were shown previously to be correctly perceived as belonging to either the solid or aggregate material typology [15,16,18].

The six signals had different features in terms of duration, amplitude, temporal evolution, and spectrum (see Fig. 1). Since the audio frequency range, viz. 20 Hz to 20 kHz, is far wider than the vibrotactile frequency range, viz. 10 Hz to 1.0 kHz, in order to simulate at haptic level the six materials the audio signals were converted into vibrotactile signals by means of spectrum truncation and pitch shifting. The amount of shift was designed to preserve the structural features of the original signals simulating the six materials.

During the experiment, participants were presented once with each of the possible audio–haptic combinations of the six materials, for a total of 36 trials (i.e., 30 audio–haptic combinations of two different materials and 6 of the same material in the audio and haptic modalities).

On each trial of the A – H (NI) and A – H (I) conditions (A = audio; H = haptic, NI = non-interactive, I = interactive), participants were presented with the auditory stimulus first and with the haptic stimulus second. On each trial of the H – A (NI) and H – A (I) conditions, participants were presented with the haptic stimulus first and with the auditory stimulus second. On each of these four conditions, each trial lasted 10 s, 5 s for both the auditory and haptic stimuli, which were presented one immediately after the other. In presence of the haptic stimuli participants were also provided with a continuous 60 dB SPL pink noise over the headphones in order to mask the audible output generated by the haptic shoes as result of the activation of the actuators.

On each trial of the A + H (NI) and A + H (I) conditions, the auditory and haptic stimuli were presented simultaneously (trial duration = 10 s). On each trial of the non-interactive conditions, participants were presented with 12 simulated footsteps (6 for each of the two modalities during the A – H (NI) and H – A (NI) conditions).

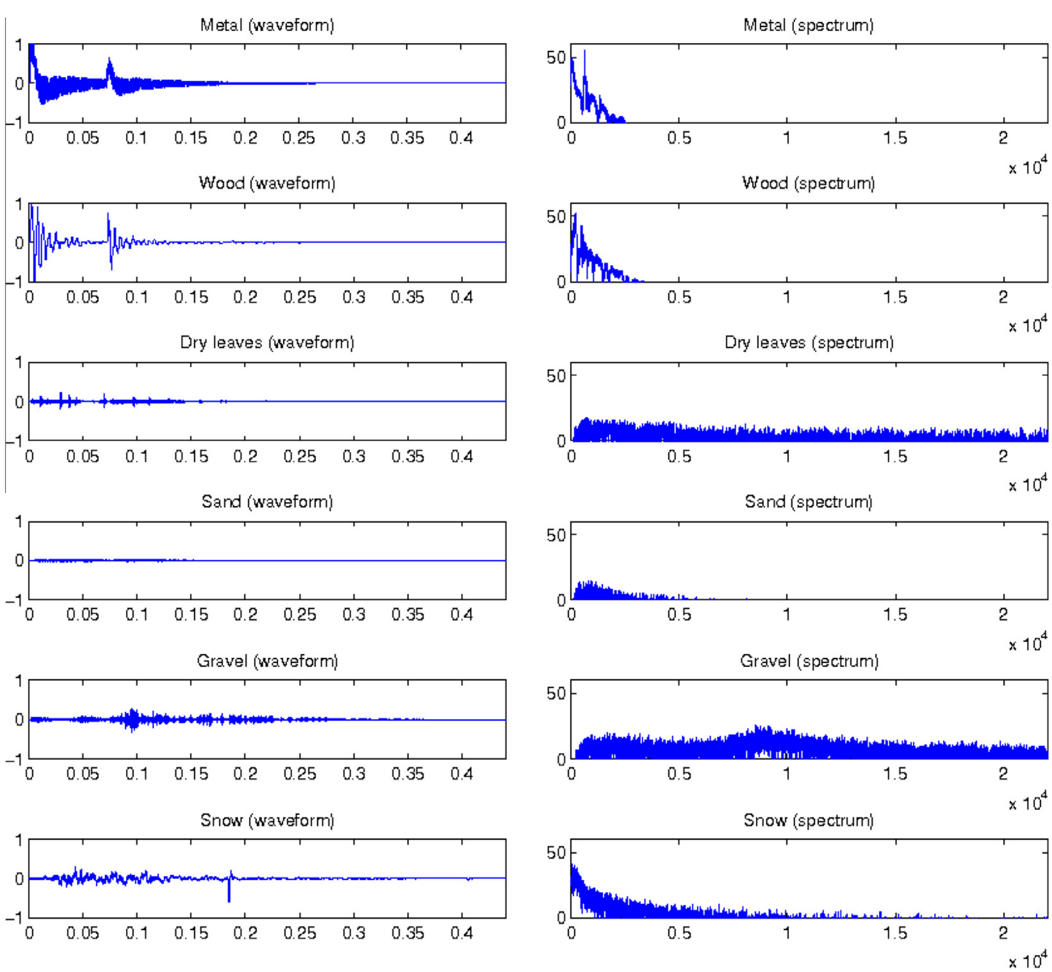

Fig. 1. Typical waveforms (left) and spectra (right) of the six simulated materials. The time axis of the waveforms is in seconds, the magnitude of the spectra is in decibel.

2.3. Procedure

Participants were asked to evaluate the degree of coherence between the auditory and haptic stimuli. In particular, during conditions A − H (NI) and A − H (I), they were instructed to "Rate the extent to which the haptic stimulus is coherent with the auditory stimulus"; during conditions H − A (NI) and H − A (I), they were instructed to "Rate the extent to which the auditory stimulus is coherent with the haptic stimulus"; during conditions A + H (NI) and A + H (I), they were instructed to "Rate the extent to which the auditory and haptic stimuli are coherent with each other". Ratings were given on a 9-point Likert scale (1 = very low coherence, 9 = very high coherence).

Participants were allowed to experience each configuration as much as they wanted before giving an answer. Before performing the task, they were presented with two practice trials in order to get familiar with the system. To this purpose, the creaking wood and frozen snow materials coherently matching at auditory and haptic level were chosen. These two materials were not among those involved in the experiment.

Participants were never informed of which material was simulated by the different experimental stimuli.

The experiment lasted, on average, 10 and 13 min in the non-interactive and interactive conditions, respectively.

2.4. Hypotheses

Our hypotheses were manifold. First of all, based on the results reported in [14,16] and [18], we hypothesized higher ratings for the audio–haptic stimuli belonging to the same typology rather than those of different typology. Secondly, since the auditory and haptic synthesis were driven by signals coincident for nature, duration and temporal evolution, as well as similar in amplitude and spectrum, we expected that the audio–haptic pairs of the same material would have been evaluated with the highest scores. Thirdly, we hypothesized an increment of the matching ability in conditions A + H (NI) and A + H (I) rather than in conditions A − H (NI) and H − A (NI), and A − H (I) and H − A (I) since participants would have had to keep in memory the auditory and the haptic signals in order to compare them, rather than making the comparison when the two signals were provided at the same time. In addition no differences between results of conditions A − H (NI) and H − A (NI) on the one hand, and A − H (I) and H − A (I) on the other hand were expected. Finally, we expected that the introduction of the interactivity would have improved the participants' matching ability rather than those in the non-interactive conditions.

2.5. Participants

Ninety individuals (57 M, 44 F, mean age = 24.23, SD = 4.01) participated in the experiment. An equal number of individuals (15) was assigned to each of the six experimental conditions. All participants reported normal hearing conditions and no locomotion problems. The size of the utilized pair of sandals was 43 (EUR). In order for the size of sandals not to affect performance, subjects wore shoes sizes from 41 to 45.

3. Results

Results are illustrated in Tables 1 and 2. In such tables the cells corresponding to average scores for the pairs of audio–haptic/haptic–audio stimuli of the same typology (i.e., solid–solid and aggregate–aggregate) are highlighted with gray, and the scores of the stimuli with the same material presented at auditory and haptic level are typed in bold. Statistical analysis was performed on the collected data by means of ANOVAs with repeated measures. All post hoc analyses were performed using the Tukey's procedure (p-value was set at a significant $p < 0.05$).

First of all, ANOVAs were performed on the data collected from the three non-interactive experiments and for the three interactive ones, by considering the four typologies for the response dependent variable (see Fig. 2). Concerning the non-interactive case, the ANOVA yielded a significant main effect for the four pairs of typologies, $F(3,132) = 52.43$, $p < 0.001$. The pair wise comparison revealed that the average score of the pair of typologies aggregate–aggregate was significantly higher than that of the pairs of typologies aggregate–solid and solid–aggregate, both $p < 0.001$. The same behavior was noticed for the average score of the pair solid–solid versus that of the pairs aggregate–solid and solid-aggregate (both $p < 0.001$). As

Table 1
Results of the non-interactive conditions. The cells corresponding to average scores for the pairs of audio–haptic/haptic–audio stimuli of the same typology are highlighted with gray, and the scores of the stimuli with the same material presented at auditory and haptic level are typed in bold.

Haptic stimulus	Auditory stimulus					
	Wood	Metal	Snow	Gravel	Sand	Dry leaves
(a) Results of condition A − H (NI)						
Wood	**6.73**	6.8	4.66	4.73	5.66	4.86
Metal	6.6	**6.53**	1.93	3.26	4.13	3.6
Snow	3.8	1.93	**7.93**	6.13	6.93	5.13
Gravel	2.86	2.46	6.53	**6.26**	7.53	6.46
Sand	4	3.93	5.6	6.33	**7.13**	6.6
Dry leaves	4.06	2.8	5.13	6.2	6.53	**6.93**
(b) Results of condition H − A (NI)						
Wood	**5.33**	4.86	3.53	3.86	4.13	4.6
Metal	5	**5.8**	2.2	2.13	2.8	2.8
Snow	5.53	2.86	**8.2**	5.73	5.6	5.4
Gravel	5.33	4.06	5.73	**5.66**	5.46	5.2
Sand	5.26	5.33	5	7.33	**6.73**	7.2
Dry leaves	5.33	5.06	5.53	5.56	5.73	**6.53**
(c) Results of condition A + H (NI)						
Wood	**6.8**	7	4.53	4.8	5.53	6.06
Metal	7	**6.06**	3.2	5.86	5.73	5.6
Snow	3.6	2.26	**8.26**	6.73	6.6	6.6
Gravel	3.53	3	6.86	**6.73**	8.26	7.86
Sand	5	3.8	6.53	6.53	**7.93**	7.8
Dry leaves	4.4	4.13	6.06	6.66	7.86	**6.86**

Table 2
Results of the interactive conditions. The cells corresponding to average scores for the pairs of audio–haptic/haptic–audio stimuli of the same typology are highlighted with gray, and the scores of the stimuli with the same material presented at auditory and haptic level are typed in bold.

Haptic stimulus	Auditory stimulus					
	Wood	Metal	Snow	Gravel	Sand	Dry leaves
(a) Results of condition A − H (I)						
Wood	**7.06**	4.4	5.06	6.2	6.73	7.06
Metal	5.86	**4.06**	6.33	6.33	7.8	6.06
Snow	4.8	2.93	**7.33**	6.6	6.2	6.8
Gravel	4.53	3.4	6.86	**5.86**	7.33	6.93
Sand	3.93	4	6.73	6.13	**6.53**	6.73
Dry leaves	5	3.86	4.93	5.33	5.73	**6.2**
(b) Results of condition H − A (I)						
Wood	**4.46**	3.73	3	4.13	3.4	4.73
Metal	3.4	**2.73**	2.33	3	2.6	3.33
Snow	5.4	5.86	**7.6**	6.06	6.13	3.46
Gravel	6.53	5.46	6.46	**4.26**	5.6	5.06
Sand	6.33	7.46	5.93	6.66	**6.93**	6.46
Dry leaves	7.2	6.2	5.06	5.73	6.6	**5.8**
(c) Results of condition A + H (I)						
Wood	**5.8**	6.93	6.2	6.8	6.4	7.2
Metal	5.33	**5.86**	5.66	6.46	7	6.93
Snow	5.33	4.33	**7.8**	6.66	6.66	7
Gravel	5.33	5.46	5.06	**5.26**	6.8	7.06
Sand	4.86	4.53	6.06	5.93	**6.4**	7.13
Dry leaves	4.86	4.46	4.2	4.53	6.13	**6.46**

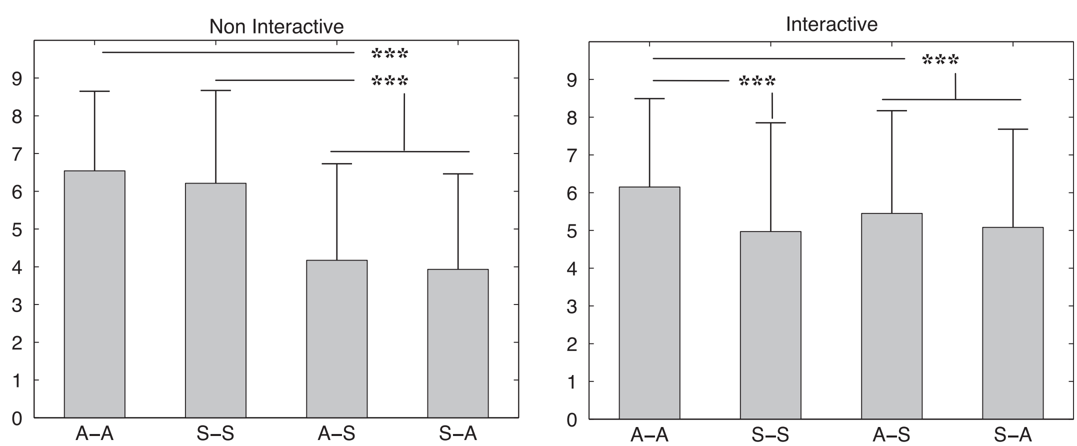

Fig. 2. Graphical representation of the mean and the standard deviation for participants' rankings corresponding to the four pairs of material typologies in the non-interactive (left) and interactive (right) conditions. Legend: A–A represents the audio–haptic pairs of aggregate–aggregate typologies, S–S the solid–solid, A–S the aggregate–solid, and S–A the solid–aggregate; *** represents $p < 0.001$.

regards the interactive case, the ANOVA showed a significant main effect for the four pairs of typologies, $F(3,132) = 6.865$, $p < 0.001$. The pair wise comparison revealed that the average score of the pair of typologies aggregate–aggregate was significantly higher not only than that of the pairs of typologies aggregate–solid and solid–aggregate (both $p < 0.001$), but also than the pair solid–solid ($p < 0.001$).

Further ANOVAs were performed by considering the audio–haptic pairs with same material belonging to the aggregate (e.g., snow–snow), as well as to the solid (e.g., wood–wood), and those with different materials (e.g., wood–snow) for the response dependent variable (see Fig. 3). Regarding the non-interactive case, the ANOVA showed a significant main effect for the four pairs of typologies, $F(2,88) = 52.43$, $p < 0.001$. The pair wise comparison revealed that the average score of the audio–haptic pairs with same material belonging to the aggregate typology was significantly higher than that of both the pairs with materials belonging to different typologies ($p < 0.001$), and of the pairs with same material belonging to the solid typology ($p < 0.05$). In addition, the average score of the audio–haptic pairs with same material belonging to the solid typology was significantly higher than that of the pairs with materials belonging to different typologies ($p < 0.001$). As for the interactive case, the ANOVA yielded a significant main effect for the four pairs of typologies, $F(2,88) = 10.49$, $p < 0.001$. The pair wise comparison revealed that the average score of the audio–haptic pairs with same material belonging to the aggregate typology was significantly higher than that of both the pairs with materials belonging to different typologies ($p < 0.001$), and of the pairs with same material belonging to the solid typology ($p < 0.001$).

Moreover, ANOVAs were conducted for each experimental condition separately to test the significance of differences both between the four typologies, and between the pairs of same material and the pairs with different materials.

Results of condition A − H (NI) are illustrated in Table 1(a). The first noticeable thing emerging from results is that the average scores corresponding to the stimuli having the same material

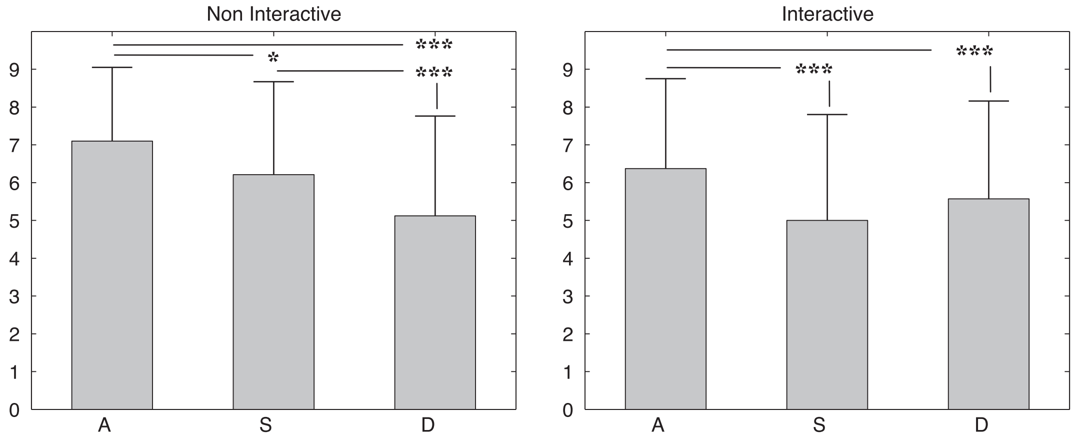

Fig. 3. Graphical representation of the mean and the standard deviation for participants' rankings corresponding to the audio–haptic pairs with same and different material in the non-interactive (left) and interactive (right) conditions. Legend: A represents the audio–haptic pairs in the aggregate–aggregate typologies, S in the solid–solid, D those belonging to the aggregate–solid and solid–aggregate; * represents $p < 0.05$, and *** represents $p < 0.001$.

typology presented both at auditory and haptic level are generally higher than the average scores corresponding to the stimuli not within the same material typology. The repeated measure ANOVA showed a significant main effect for the four pairs of typologies, $F(3,42) = 31.9$, $p < 0.001$. The pair wise comparison revealed that the average score of the pair of typologies aggregate–aggregate was significantly higher than that of the pairs of typologies aggregate–solid and solid–aggregate, both $p < 0.001$. The same behavior was noticed for the average score of the pair solid–solid versus that of the pairs aggregate-solid and solid-aggregate (both $p < 0.001$).

Considering the audio–haptic pairs with same material it is possible to notice that the evaluations are among the highest in the table. The ANOVA performed between the audio–haptic pairs with same material and all he others revealed a significant main effect $F(1,14) = 34.322$, $p < 0.001$. In addition, the average score of the audio–haptic pairs with same material was also proven to be significantly higher than that of the pairs belonging to the same typology $F(1,14) = 4.974$, $p = 0.042$.

Results of condition H − A (NI) are illustrated in Table 1(b). Similarly to the results of condition 1, the scores corresponding to the stimuli having the same material typology presented both at auditory and haptic level are most of the times higher than the average scores corresponding to the stimuli not within the same material typology. In particular, the ANOVA showed a significant main effect for the four pairs of typologies, $F(3,42) = 17.525$, $p < 0.001$. The post hoc analysis showed that the average score of the pair aggregate–aggregate was significantly higher than that of solid-aggregate and aggregate–solid, both $p < 0.001$; likewise for the average score of the pair solid–solid versus typology aggregate–solid ($p < x0.001$).

The evaluations of the audio–haptic pairs with same material were among the highest in the table. The ANOVA performed between the audio–haptic pairs with same material and all the others revealed a significant main effect $F(1,14) = 22.45$, $p < 0.001$. In addition, the average score of the audio–haptic pairs with same material was also proven to be significantly higher than that of the pairs belonging to the same typology $F(1,14) = 6.668$, $p = 0.021$.

Table 1(c) illustrates the results of condition A + H (NI). Also in this case, it is possible to notice how the average scores corresponding to the stimuli within the same material typology presented both at auditory and haptic level are generally higher than the average scores corresponding to the stimuli not within the same typology. In more detail, the ANOVA performed between the four different pairs of typologies revealed significant main effect, $F(3,42) = 24.669$, $p < 0.001$. The post hoc analysis showed that the average score of the pair aggregate–aggregate was significantly higher than that of the pairs aggregate–solid and solid–aggregate, both $p < 0.001$. The same behavior was noticed for the average score of the pairs solid–solid versus aggregate–solid and solid–aggregate (both $p < 0.001$).

The evaluations of the audio–haptic pairs with same material were among the highest in the table. The ANOVA performed between the audio–haptic pairs with same material and all the others revealed a significant main effect $F(1,14) = 16.89$, $p = 0.001$. However, the audio–haptic pairs with same material were not proven to be significantly higher than the pairs belonging to the same typology.

Table 2(a) illustrates the results for the interactive condition A − H (I). The ANOVA showed a significant main effect for the four pairs of typologies, $F(3,42) = 18.058$, $p < 0.001$. The pair wise comparison revealed that the average score of the pair aggregate–aggregate was significantly higher than that of the pair solid-aggregate ($p < 0.001$), but not than the pair of typologies aggregate–solid. Similarly, the average score of the pair solid–solid was significantly higher than that of the pair solid–aggregate ($p = 0.002$), but significantly lower than that of typology aggregate–solid ($p = 0.013$). In addition the average score of the pair solid–solid was significantly lower than that of the pair aggregate–aggregate ($p = 0.009$).

The evaluations of the audio–haptic pairs with same material were among the highest in the table. The average value for such audio–haptic pairs was higher than that of all the others pairs but not in a significant way. Analogously the average score of the audio–haptic pairs with same material was not proven to be significantly higher than that of the pairs belonging to the same typology.

Results of condition H − A (I) are illustrated in Table 2(b). The ANOVA performed between the four different typologies turned out to be significant, $F(3,42) = 21.073$, $p < 0.001$. The post hoc analysis showed that the average score of the pair aggregate–aggregate was significantly higher than that of the pair solid-aggregate, $p < 0.001$, but not than that of typology aggregate–solid. As concerns the average score of the typology solid–solid, it was significantly higher than that of typology solid-aggregate ($p < 0.001$), but lower than that of typology aggregate–solid, even if not in a significant way. In addition the average score of the pair solid–solid was significantly lower than that of the pair aggregate–aggregate ($p < 0.001$).

The evaluations of the audio–haptic pairs with same material were among the highest in the table. The average value for such audio–haptic pairs was not proven to be significantly higher than that of all the others pairs nor than that of the pairs belonging to the same typology.

Table 2(c) illustrates the results for the interactive condition A + H (I), which consisted of the simultaneous presentation of the auditory and haptic stimulus, and participants were asked to evaluate to which extent the auditory and the haptic stimuli were coherent with each other. The ANOVA showed a significant main effect for the four pairs of typologies, $F(3,42) = 6.64$, $p < 0.001$. The post hoc analysis showed that the average score of the pair aggregate–aggregate was significantly higher than that of typology solid-aggregate ($p < 0.001$), but lower than that of typology aggregate–solid, even if not in a significant way. Concerning the average score of the pair solid–solid, it was higher than that of typology solid–aggregate ($p = 0.01$), but lower than that of typology aggregate–solid, even if not in a significant way. The evaluations of the audio–haptic pairs with same material were among the highest in the table. However, the average value for such audio–haptic pairs was not proven to be significantly higher than that of all the others pairs nor than that of the pairs belonging to the same typology.

The average times taken by participants to complete the assigned tasks were about three minutes smaller for the non-interactive conditions compared to the interactive ones. A *t*-test proved the statistical significance of such difference ($t(67.762) = -5.0988$, $p < 0.001$).

4. Discussion

As illustrated in Fig. 2, significantly higher evaluations were found in the non-interactive conditions for the typologies aggregate–aggregate versus aggregate–solid and solid–aggregate, as well as for solid–solid versus aggregate–solid and solid–aggregate. Consequently these findings confirm our hypothesis of higher evaluations for the audio–haptic stimuli belonging to the same typology rather than those of the different typologies, and in addition are in line with the trend found in the results illustrated in [16], in which participants were capable of discriminating both at auditory and haptic level the typology of the presented materials. Conversely, such hypothesis was only partially confirmed for the interactive conditions since significantly higher ratings were found only for aggregate–aggregate versus solid–aggregate and aggregate–solid. In addition, looking at the gray cells of the Table 2(a), (b) and (c), it is possible to notice how the average scores inside the same typologies were in most of the cases quite homogeneous, indicating that the matching of the materials belonging to the same typology was not too accurate. Thus, such result partially confirms the trend found in the results illustrated in [18], in which participants were capable of discriminating both at auditory and haptic level the typology of the presented materials.

The hypothesis that the audio–haptic pairs of same material would receive the highest coherence scores, was confirmed for the non-interactive conditions and partially confirmed for the interactive ones, as illustrated in Fig. 3. Indeed, in the non-interactive conditions the evaluations of the audio–haptic pairs of same material belonging both to the aggregate and solid typology were proven to be significantly higher than the pairs with different materials, while for the interactive conditions significance was found only for the pairs of same material belonging to the aggregate typology.

A deeper comparison between Table 1(a), (b) and (c), and Table 2(a), (b) and (c), as well as between the two plots of Fig. 2, reveals on the one hand that the evaluations of the audio–haptic stimuli of the same typology were on average lower for the interactive conditions compared to the non-interactive ones, and on the other hand that the evaluations of the stimuli of different typology were on average higher for the interactive conditions compared to the non-interactive ones.

In addition, looking at the grey cells of the Table 2(a), (b) and (c), it is possible to notice how the average scores inside the same typologies were in most of the cases quite homogeneous, indicating that the matching of the materials belonging to the same typology was not too accurate in the interactive conditions.

Furthermore, the average times taken by participants to complete the assigned tasks were significantly larger for the interactive conditions compared to the non-interactive ones, although the actual duration of the trials was identical in the two set of conditions. This finds an explanation in the fact that participants, who were allowed to try each trial as much as they wanted before giving an answer, repeated the trials more times during the interactive condition.

All these details lead to the conclusion that subjects performed better the task non-interactively rather than interactively, thus contradicting our initial hypothesis. First of all, this might be due to the fact that during the non-interactive conditions subjects could concentrate more on the task itself instead of focusing also on the act of walking. Indeed, on the one hand it has been proved that when walking on a compliant surface the central nervous system controls the whole body in order to maximize stability [26]. On the other hand it has been shown that the use of plantar cutaneous vibration feedback is sufficient to elicit a percept of compliance during walking [13]. Therefore it is possible that participants, especially in presence of an aggregate surface at haptic level, were confused or distracted from their task in order to compensate for the change in compliance induced by the haptic feedback. Secondly, during conditions A − H (I) and H − A (I) a haptic sensation was present also while walking with only auditory feedback (and no haptic feedback). This situation differs from the one of the non-interactive conditions in which the haptic sensation arose clearly only when provided. However, similar results were found also in condition A + H (I), therefore it is more plausible to think that the difference between the two sets of conditions was mainly due to the limits of the simulations. Indeed, a probable explanation for these findings lies in the fact that participants actually walked on a surface that was already solid (carpet on a concrete floor), and thus the real vibrations arriving to the feet when hitting the ground summed to the simulated ones.

One of the biggest differences between the simulations of solid and aggregate materials is their temporal duration (see Fig. 1). Simulations of solid materials are characterized by a short duration, while the aggregate materials are simulated using signals having a longer temporal evolution. While walking the feet are in contact with the ground for a time longer than the duration of the haptic simulation of the solid surface therefore the incongruence between a solid surface provided at haptic level and an aggregate surface provide at auditory level could be less noticeable. Therefore such results confirm that using the proposed shoe system it is indeed a hard task to enhance, at haptic level, a realistic haptic sensation of walking over a solid surface while actually walking over a surface that is already solid.

Conversely, in presence of a simulated aggregate surface interactively provided, the contribution arriving to the feet is the sum of the simulated vibrations and those coming from the ground (solid surface in the case of our experiment), and therefore the incongruence between an aggregate surface provided at haptic level and a solid surface provided at auditory level could be less noticeable.

Finally, no substantial differences were found either between results of the three non-interactive conditions or between those of the interactive ones. Participants were no more accurate in matching haptic and audio stimuli when presented sequentially, regardless of presentation order, or when presented simultaneously.

5. Conclusions

In this paper, we described an experiment whose goal was to assess the ability of subjects in matching pairs of simulated materials presented at auditory and haptic level both interactively and non-interactively. Better matching ability was found for the non-interactive case compared to the interactive one, although this may be due to the limits of the technology used for the haptic simulations, especially in what concerns the solid surfaces. Overall, subjects expressed a higher level of semantic congruence for those materials which belonged to the same typology over that for materials belonging to a different typology. These results, together with those reported in [14,16] and [18] show that material typology is processed very consistently in both the auditory and haptic modalities. This suggests that there are similarities between the perceptual spaces of the two modalities, and this could be explained by the presence of a shared representation underlying both modalities. Further studies should be performed to verify such hypothesis.

Finally, these results are important in order to provide guidelines to designers of audio–haptic interfaces for navigation in virtual environments. Indeed, practical indications can be derived from this study. Results of condition A + H (NI) reveal that when simulating non-interactively the audio–haptic walk on a surface material, in presence of a footstep sound on a solid material it is possible to utilize at haptic level any other solid material, but not an aggregate one, which would be perceived as not fitting well; similarly when providing the sound of a footstep on an aggregate material it is possible to use at haptic level any other aggregate material, but not a solid one. Conversely results of condition A + H (I) only indicate the necessity of not providing an aggregate material at haptic level when at auditory level a solid surface is presented.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme under FET-Open grant agreement 222107 NIW – Natural Interactive Walking, and from the Danish Council for Independent Research – Technology and Production Sciences (FTP), Grant No. 12–131985.

References

[1] Spence C. Audiovisual multisensory integration. Acoust Sci Technol 2007;28(2):61–70.
[2] Molholm S, Ritter W, Javitt D, Foxe J. Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. Cerebral Cortex 2004;14(4):452.
[3] Laurienti P, Kraft R, Maldjian J, Burdette J, Wallace M. Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 2004;158(4):405–14.
[4] Chen Y, Spence C. When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. Cognition 2010;114(3):389–404.
[5] Koppen C, Alsius A, Spence C. Semantic congruency and the colavita visual dominance effect. Exp Brain Res 2008;184(4):533–46.
[6] Lederman S et al. Auditory texture perception. Perception 1979;8(1):93–103.
[7] Guest S, Catmur C, Lloyd D, Spence C. Audiotactile interactions in roughness perception. Exp Brain Res 2002;146:161–71.
[8] DiFranco D, Beauregard G, Srinivasan M. The effect of auditory cues on the haptic perception of stiffness in virtual environments. in: Proceedings of the ASME dynamic systems and control division, vol. 61; 1997. p. 17–22.
[9] Avanzini F, Crosato P. Haptic-auditory rendering and perception of contact stiffness. Lect Notes Comput Sci 2006;4129:24–35.
[10] Li XF, Logan RJ, Pastore RE. Perception of acoustic source characteristics: walking sounds. J Acoust Soc Am 1991;90(6):3036–49.
[11] Pastore R, Flint J, Gaston J, Solomon M. Auditory event perception: the source–perception loop for posture in human gait. Attention, Perception, Psychophys 2008;70(1):13–29.
[12] Kobayashi Y, Osaka R, Hara T, Fujimoto H. How accurately people can discriminate the differences of floor materials with various elasticities. IEEE Trans Neural Rehab Syst Eng 2008;16(1):99–105.
[13] Visell Y, Giordano B, Millet G, Cooperstock J. Vibration influences haptic perception of surface compliance during walking. PLoS ONE 2011;6(3):11.
[14] Giordano B, Visell Y, Yao H-Y, Hayward V, Cooperstock J, McAdams S. Identification of walked-upon materials in auditory, kinesthetic, haptic and audio–haptic conditions. J Acoust Soc Am 2012;131:4002–12.
[15] Nordahl R, Serafin S, Turchet L. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: Proceedings of the IEEE virtual reality conference. IEEE Press; 2010. p. 147–53.
[16] Nordahl R, Berrezag A, Dimitrov S, Turchet L, Hayward V, Serafin S. Preliminary experiment combining virtual reality haptic shoes and audio synthesis. In: Haptics: generating and perceiving tangible sensations. Lecture notes in computer science, vol. 6192. Berlin Heidelberg: Springer; 2010. p. 123–9.
[17] Turchet L, Burelli P, Serafin S. Haptic feedback for enhancing realism of walking simulations. IEEE Trans Haptics 2013;6(1):35–45.
[18] Serafin S, Turchet L, Nordahl R, Dimitrov S, Berrezag A, Hayward V. Identification of virtual grounds using virtual reality haptic shoes and sound synthesis. in: Proceedings of eurohaptics symposium on haptic and audio-visual stimuli: enhancing experiences and interaction; 2010. p. 61–70.
[19] Turchet L, Serafin S, Dimitrov S, Nordahl R. Physically based sound synthesis and control of footsteps sounds. in: Proceedings of digital audio effects conference; 2010. p. 161–168.
[20] Turchet L, Nordahl R, Berrezag A, Dimitrov S, Hayward V, Serafin S. Audio–haptic physically based simulation of walking on different grounds. In: Proceedings of IEEE international workshop on multimedia signal processing. IEEE Press; 2010. p. 269–73.
[21] Yao H-Y, Hayward V. Design and analysis of a recoil-type vibrotactile transducer. J Acoust Soc Am 2010;128(2):619–27.
[22] Avanzini F, Rocchesso D. Modeling collision sounds: Non-linear contact force. in: Proceedings of digital audio effects conference; 2001. p. 61–66.
[23] Cook P. Physically informed sonic modeling (phism): synthesis of percussive sounds. Comput Music J 1997;21(3):38–49.
[24] Hunt KH, Crossley FRE. Coefficient of restitution interpreted as damping in vibroimpact. ASME J Appl Mech 1975;42(2):440–5.
[25] Turchet L, Serafin S. An investigation on temporal aspects in the audio–haptic simulation of footsteps. In: Multidisciplinary aspects of time and time perception. Lecture notes in computer science, vol. 6789. Berlin Heidelberg: Springer; 2011. p. 101–15.
[26] MacLellan MJ, Patla AE. Adaptations of walking pattern on a compliant surface to regulate dynamic stability. Exp Brain Res 2006;173(3):521–30.