PAPER

Temporal and amplitude aspects in sonically simulating walking over bumps, holes and flat surfaces

Luca Turchet* and Stefania Serafin†

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

(*Received 21 June 2011, Accepted for publication 9 March 2012*)

Abstract: This paper describes an experiment whose goal is to assess the role of temporal and amplitude variations in sonically simulating the act of walking over a bump or a hole. In particular, it has been investigated whether the timing between heel and toe and the timing between footsteps, as well as variations in the amplitude of heel and toe affect the perception of walking on unflat surfaces. Forty five subjects participated to three between-subjects experiments where they were asked to interact with a desktop system simulating bumps, holes and flat surfaces by means of auditory cues. Results show that it is possible to simulate a bump or a hole by only using temporal information in the auditory modality. Furthermore results show that the proposed amplitude variations are not sufficient to provide the information concerning uneven surfaces.

Keywords: Footsteps, Bumps, Holes, Amplitude

PACS number: 43.66.Jh, 43.66.Lj, 43.66.MK [doi:10.1250/ast.33.291]

1. INTRODUCTION

Footstep sounds represent important elements in multimedia productions such as movies and videogames. Chion *et al.* write of footstep sounds as being rich in what he refers to as *materializing sound indices* — those features that can lend concreteness and materiality to what is on-screen, or contrarily, make it seem abstracted and unreal [1].

In the auditory perception and sound design and synthesis community, sonic properties of footsteps have been extensively investigated. From the perceptual point of view, previous research has shown that it is possible to recognize the gender of a human walker only by listening to recorded footsteps [2]. Moreover, footstep sounds of a person walking on a wooden floor provide information about the gender, age, size, and emotional intention of the person, and hardness and material of both shoes and floor [3]. Other studies demonstrated the possibility of recognizing simulated surfaces subjects are walking upon [4], as well as emotional intentions of the walker [5,6].

To our knowledge, all previous research on walking sounds from a simulation point of view has focused on the act of walking on flat surfaces [7–11].

On the other hand, the human-computer interaction community has explored the possibility of simulating bumps and holes, by implementing a so-called pseudo-haptic simulation [12]. The main idea of such research was to investigate whether it is possible to simulate a bump or a hole on a screen by only using visual feedback. This illusion was achieved by creating a visual interface where the control-display ratio between the motion of the mouse and the cursor is not linear. In particular, when simulating a bump, the cursor on the screen was decelerating until reaching the top of the bump and then accelerating, while when simulating a hole the cursor first accelerated and then decelerated. Experiments showed that subjects could successfully recognize a hole or a bump with this system [12].

Such research has recently been extended in [13], where the authors investigated whether it is possible to simulate the illusion of walking on a hole or a bump only by using visual feedback. Three parameters were considered in such simulation: orientation, velocity and height, and their combination. The experiments were run both actively, having users wear an head mounted display, as well as passively, having users interact with a desktop simulation. Results showed that such visualization techniques successfully simulate bumps and holes located in the ground.

It has also been shown that it is possible to simulate a bump or a hole only by using auditory cues [14]. In this

*e-mail: tur@create.aau.dk
†e-mail: sts@create.aau.dk

case, the auditory cues which create the sensation of walking on unflat surfaces are given by varying the temporal distance between footsteps, as well as the distance between heel and toe events in single footsteps.

Moreover, this work was extended by implementing a multimodal (audio-visual) simulation of walking on a bump or a hole [15]. Results in this case show that the auditory cues reinforce the visual cues when coherent cues are provided in both modalities. When subjects are exposed to conflicting cues, for example by simulating visually the act of walking on a bump and auditory the act of walking on a hole, usually the visual cues are dominant, apart from when the velocity effect is the visual parameter varied. This might be due to the higher temporal resolution of the auditory system versus the visual system.

In this paper, we are interested in extending the possibilities offered by implementing such pseudo-haptic feedback from the sonic point of view. Where most research until now has focused on the person walking, in this experiment we are instead addressing the significance of the surface the person is stepping upon. This is achieved by investigating whether it is possible to simulate the act of walking on unflat surfaces by using both temporal variations among footsteps as well as variations in amplitude. The reason to investigate temporal and amplitude aspects of our simulations is that the parameters of the extracted ground reaction force (GRF), described in the following section, are precisely temporal and amplitude variations. So our goal is to investigate whether modifications to the GRF create the perception of walking on unflat surfaces.

We first describe the sound synthesis engine used, and we then propose three experiments which assess the role of temporal and amplitude variations in simulating the act of walking on unflat surfaces.

2. SYNTHESIS OF FOOTSTEP SOUNDS

In previous research, we proposed a sound synthesis engine able to simulate footstep sounds on aggregate and solid surfaces [16]. Such engine is based on physical models which are driven by a signal, in the audio domain, expressing the ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. In our simulations the GRF corresponds to the amplitude envelope extracted from an audio signal containing a footstep sound. The engine can work both offline and in real-time. The two approaches differ for the way the input GRF is generated. Concerning the real-time implementation, various systems for the generation of such input have been developed and tested [4,16–19]. During the offline work, the input signal is not detected in real-time, but it consists of an audio file from which the GRF is extracted. Such file consists of a recording of a person walking on a real surface. Better results in terms of the GRF detection can be found on audio recordings of walking on solid surfaces and with a small amount of background noise. The different envelope profiles of each step in the file are extracted and fed to the engine which produces the synthesized footstep sounds according to the choice of the surface to be simulated.

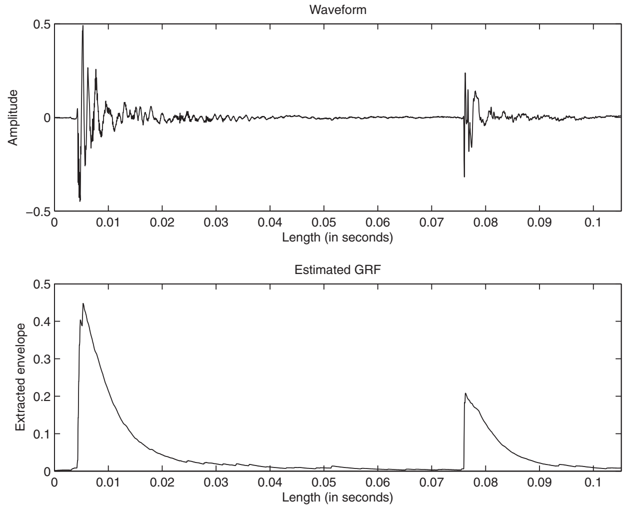

Fig. 1 Waveform of the used footstep on concrete (top) and relative extracted GRF (bottom).

In this particular set of experiments, we adopted the offline use of the engine. To control the engine, we created different audio files placing at various temporal patterns the recording of an unique real footstep sound on concrete. Such sound was chosen among those available in the Hollywood Edge sound effects library.* As an example, Fig. 1 shows the waveform of the chosen footstep on concrete on top, and its corresponding GRF on the bottom. Three types of surface profiles have been created starting from the footstep sound generator: bumps, holes, and flat surfaces (see Fig. 2). The techniques adopted to render them at auditory level are illustrated in Sections 4, 5, 6.

For the purpose of these experiments, two types of surfaces, one solid (wood) and one aggregate (gravel), were chosen. The reason for choosing two materials was to assess whether the surface type affected the quality of the results.

3. EXPERIMENT DESIGN

We conducted three between-subjects experiments whose goal was to investigate the ability of subjects to recognize if the sounds they were exposed to corresponded to walking on a bump, a hole or a flat surface:

(1) Experiment 1: recognition of bumps, holes and flat surfaces by means of temporal intervals variations

*Hollywood Edge sound effects library: www.hollywoodedge.com

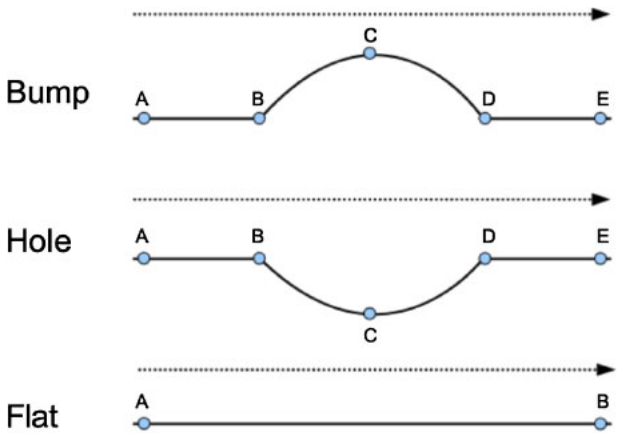

Fig. 2 The three types of surfaces modelled.

(2) Experiment 2: recognition of bumps, holes and flat surfaces by means of amplitude variations
(3) Experiment 3: recognition of bumps, holes and flat surfaces by means of combinations of amplitude and temporal intervals variations

The between-subjects design was preferred to a within-subject design in order to avoid a possible learning effect.

During experiment 1 the audio files were created placing at different temporal intervals the footstep sound generator, keeping unvaried their amplitude. The base idea is that in a real environment, a person generally walks slower on ascending slopes, and faster on descending slopes. We transposed this information in our experiment by modifying the time intervals both between footsteps and between the heel and toe information in each footstep (see Section 4). Conversely in experiment 2 the bumps and holes were rendered by means of amplitude variations, keeping fixed the time intervals both between footsteps and between heel and toe (see Section 5). The chosen mapping strategy was to progressively decrease the footstep sounds amplitude during ascending slopes and gradually increase it during descending slopes. Finally in experiment 3 the two techniques were combined (see Section 6).

In order to have the integration of the two techniques for experiment 3, the files of experiment 2 were designed in reference to those of experiment 1. In particular the same number of steps (with the same distribution according to Fig. 2) were used (see Tables 1 and 4). Moreover in oder to have consistency among experiments the same number of different surface profiles were designed (specifically 2 flat, 6 bumps and 6 holes).

One of our hypotheses was that the information provided by means of the temporal variations would have been more helpful rather the one provided by the amplitude variations. Another hypothesis was that the recognition would have improved using both the information rather the single information alone.

3.1. Participants

Forty-five participants were divided in three groups ($n = 15$) to perform the three between-subjects experiments. The three groups were composed respectively of 11 men and 4 women, aged between 20 and 29 (mean = 23.6, standard deviation = 2.84), 10 men and 5 women, aged between 20 and 29 (mean = 23.86, standard deviation = 3.04) and 10 men and 5 women, aged between 20 and 28 (mean = 23.6, standard deviation = 2.58). All participants reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment.

The participants took on average about 15, 14 and 13 minutes for experiments 1, 2 and 3 respectively.

3.2. Setup

All experiments were carried out in an acoustically isolated laboratory where the setups for the experiments were installed. They consisted of a simple graphical user interface with which the participants were asked to interact, and a spreadsheet to collect their answers. The interface was created using the Max/MSP program[†] and was composed only by buttons to be pressed. Each button was numbered, and by pressing it a sound was triggered and conveyed to the user by means of headphones. Users were asked to press each button according to their numerical order, and to write the corresponding answers on the spreadsheet.

3.3. Task

During all experiments subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface mentioned in Section 3.2. They were given the list of three different surfaces (bump, hole, flat), presented as forced alternate choice. The task consisted of recognizing to which surface the walk corresponded after the presentation of the stimulus. In addition to the classification of the surface profiles subjects were also asked to evaluate the degree of certainty of their choice on a scale from 1 to 7 (1 = very low certainty, 7 = very high certainty).

Participants were allowed to listen to the sounds as much as they wanted before giving an answer. When moving to the next stimulus they could not change the answer to the previous stimuli.

4. EXPERIMENT 1

4.1. Description of the Conditions

As mentioned in Section 3, in experiment 1 temporal intervals variations were used while any amplitude variation was involved. The amplitudes for the footstep sounds

[†]Max/MSP: www.cycling74.com

Table 1 Features of the 14 files used in experiment 1 as input to the sound engine. For a detailed description, see the text.

	Duration (in sec.)	Number of steps	Footsteps distance increment (in ms)	Footsteps distance range (in ms)	Heel-toe distance increment (in ms)	Heel-toe distance range (in ms)
flat_1	12	19	—	550 (fixed)	—	69 (fixed)
flat_2	16	19	—	750 (fixed)	—	69 (fixed)
bump_1_step	27	$31 = 4 + 12 + 11 + 4$	50	550 ⟶ 1,150	—	69 (fixed)
bump_2_step	16	$19 = 4 + 6 + 5 + 4$	100	550 ⟶ 1,150	—	69 (fixed)
hole_1_step	18	$31 = 4 + 12 + 11 + 4$	−50	750 ⟶ 150	—	69 (fixed)
hole_2_step	11	$19 = 4 + 6 + 5 + 4$	−100	750 ⟶ 150	—	69 (fixed)
bump_1_h_t	24	$31 = 4 + 12 + 11 + 4$	—	550 (fixed)	30	0 ⟶ 360 (+ 69)
bump_2_h_t	14	$19 = 4 + 6 + 5 + 4$	—	550 (fixed)	60	0 ⟶ 360 (+ 69)
hole_1_h_t	24	$31 = 4 + 12 + 11 + 4$	—	550 (fixed)	−20	240 ⟶ 0 (+ 69)
hole_2_h_t	13	$19 = 4 + 6 + 5 + 4$	—	550 (fixed)	−40	240 ⟶ 0 (+ 69)
bump_1_comb	32	$31 = 4 + 12 + 11 + 4$	50	550 ⟶ 1,150	30	0 ⟶ 360 (+ 69)
bump_2_comb	19	$19 = 4 + 6 + 5 + 4$	100	550 ⟶ 1,150	60	0 ⟶ 360 (+ 69)
hole_1_comb	22	$31 = 4 + 12 + 11 + 4$	−50	750 ⟶ 150	−20	240 ⟶ 0 (+ 69)
hole_2_comb	15	$19 = 4 + 6 + 5 + 4$	−100	750 ⟶ 150	−40	240 ⟶ 0 (+ 69)

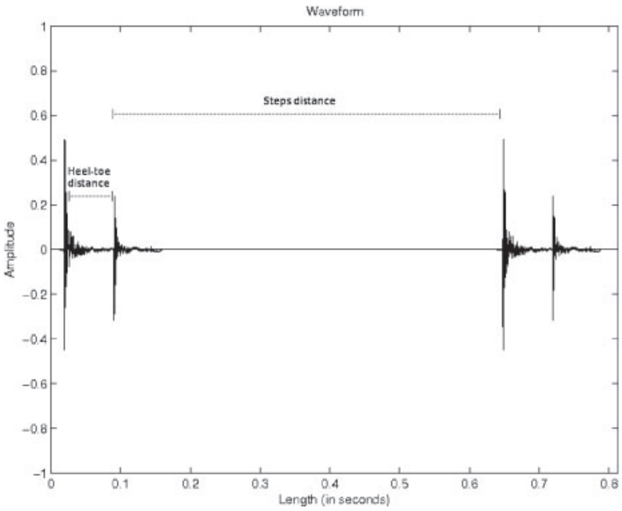

Fig. 3 Temporal distances between (named "steps distance" in the Figure), and within (named "heel-toe distance" in the Figure) footsteps.

were set to 54 dB (A) and 58 dB (A) for the wood and gravel surfaces respectively. The temporal patterns used were designed to simulate 14 different surface profiles. Specifically 2 flat, 6 bumps and 6 holes were designed. Such patterns involved three types of temporal distances. The first was the temporal distance between footsteps (i.e., the time interval between the end of the sound generated by the toe and the beginning of the sound generated by the heel of the next step), the second was the temporal distance between heel and toe (i.e., the time interval between the end of the sound generated by the heel and the beginning of the sound generated by the toe in the same step), the third consisted of the combination of the previous two (see Fig. 3).

The characteristics of the 14 files used to drive the sound engine are illustrated in Table 1. In such table the suffixes _step, _h_t and _comb indicate the type of temporal distance used for each file (footsteps distance, heel-toe distance and their combinations respectively). The equations in the "Number of steps" column indicate how the steps where placed in reference to Fig. 2. As an example, the stimulus bump_2_step was composed by 19 steps, 4 steps to go from point A to point B, 6 steps to go from point B to point C, 5 steps to go from point C to point D, and 4 steps to go from point D to point E). In order to model two different types of bumps and holes, for each category of surface modeling (by means of the three temporal distance types), two slopes where chosen. The heights of the two typologies of bumps were approximatively 5.60 and 2.80 meters. Similarly for the depth of the two modelled holes types, which were symmetrical to the bumps. These data were estimated by means of a raw calculation based on the number of steps and on the average person's stride length.

The sound engine was set in order to synthesize footstep sounds on two different kinds of materials: wood and gravel. Participants were exposed to 28 trials, where the 14 surface profiles were presented twice in randomized order (each surface profile was presented with both wood and gravel).

4.2. Results of Experiment 1

The results of experiment 1 for wood and gravel are shown in Tables 2 and 3 respectively. In both tables, the first column shows the different conditions as described in Table 1. The second, third and fourth columns illustrate the choices of the subjects (bump, hole or flat) for the different conditions they were exposed to. The fifth, sixth and seventh column report the average certainty expressed by the subjects after performing their choice; the fifth column reports the total certaintly in both correct and wrong

Table 2 Results of experiment 1 for the wood surface.

	Bump	Hole	Flat	Mean certainty			% Correct answers
				Total	Correct answers	Wrong answers	
flat_1	1		14	6	6.1429	4	93.33
flat_2		1	14	5.6667	6	1	93.33
bump_1_step	13	2		5.4286	5.5385	4.5	93.33
bump_2_step	14	1		5.4667	5.5714	4	93.33
hole_1_step	1	14		4.8	4.7857	5	93.33
hole_2_step	2	13		4.9333	5	4.5	86.66
bump_1_h_t	13	1	1	5.6	5.8469	4	86.66
bump_2_h_t	12	2	1	5.1333	5.4167	4	80
hole_1_h_t		15		4.2667	4.2667		100
hole_2_h_t	3	12		4.4	4.8182	3.25	80
bump_1_comb	14	1		5.4	5.5	4	93.33
bump_2_comb	14	1		5.1333	5.3571	2	93.33
hole_1_comb	1	14		5.1333	5.2143	4	93.33
hole_2_comb	1	14		4.9333	5	4	93.33

Table 3 Results of experiment 1 for the gravel surface.

	Bump	Hole	Flat	Mean certainty			% Correct answers
				Total	Correct answers	Wrong answers	
flat_1		3	12	5.2667	5.6667	3.6667	80
flat_2	1	1	13	5.4667	6	2	86.66
bump_1_step	12	2	1	5.8469	5.75	4.6667	80
bump_2_step	13	1	1	5.4667	5.9231	2.5	86.66
hole_1_step		14	1	5.4667	5.7857	1	93.33
hole_2_step		15		6	6		100
bump_1_h_t	14	1		5.5333	5.7857	2	93.33
bump_2_h_t	11	2	2	4.5333	4.8182	3.75	73.33
hole_1_h_t	1	12	2	4.5333	4.5833	4.3333	80
hole_2_h_t		15		4.4	4.4		100
bump_1_comb	13	2		5.3333	5.6667	4	86.66
bump_2_comb	13	1	1	5.2667	5.7692	2	86.66
hole_1_comb	2	12	1	5.2667	5.3333	5	80
hole_2_comb	2	13		5.2667	5.8469	1.5	86.66

answers, while the sixth and seventh column report the certainty in correct and uncorrect answers respectively. Finally, the last column reports the percentage of correct answers.

As the tables show, subjects could successfully recognize bumps and holes using only the auditory cues described in the previous section. In fact, as can be seen in the last column of Tables 2 and 3, the percentage of correct answers is high for all conditions, reaching also 100% of correct answers in three conditions, and with a lowest score of 73% which was reached only in one condition.

Observing columns 6 and 7, morever, it is possible to notice how subjects are quite certain when they express a correct answer. In both surfaces, indeed, the mean certainty for correct answers is always above average. On the other hand, in situations where the answer was incorrect the degree of certainty is also extremely low. This is the case, for example, in the second flat stimulus for the wood surface and the first hole stimulus in the gravel surface.

An ANOVA was performed to examine whether significant differences were present in the recognition rate among the two surfaces and among the different conditions in the same surface. Overall, no significant differences were measured in the recognition rate among the two surfaces. Moreover, no significant differences were measured in the recognition rate for the different conditions in the same material. For example, no difference was measured in the recognition rate of the first simulated bump footstep versus the second simulated bump footstep. No significant difference was furthermore measured between the recognition rate obtained when changing the temporal information between footsteps versus the one obtained when changing the temporal information within footsteps. Also, the combination of the two temporal information did not significantly enhance the recognition of a bump or a

Table 4 Features of the 14 files used in experiment 2 as input to the sound engine. For a detailed description, see the text.

	Duration (in sec.)	Number of steps	Heel/toe amplitude decrement range (in dB)	Heel/toe amplitude decrement (in dB)	Heel/toe amplitude increment range (in dB)	Heel/toe amplitude increment (in dB)
flat_1	12	19	—	—	—	—
flat_2	16	19	—	—	—	—
bump_1_t	27	$31 = 4 + 12 + 11 + 4$	heel: $0 \longrightarrow -9$	heel: 0.75	heel: $-9 \longrightarrow 9$	heel: 1.63
bump_2_t	16	$19 = 4 + 6 + 5 + 4$	heel: $0 \longrightarrow -9$	heel: 1.5	heel: $-9 \longrightarrow 9$	heel: 3.6
hole_1_t	18	$31 = 4 + 12 + 11 + 4$	heel: $9 \longrightarrow -9$	heel: 1.63	heel: $0 \longrightarrow 9$	heel: 0.75
hole_2_t	11	$19 = 4 + 6 + 5 + 4$	heel: $9 \longrightarrow -9$	heel: 3.6	heel: $0 \longrightarrow 9$	heel: 1.5
bump_1_h	24	$31 = 4 + 12 + 11 + 4$	toe: $0 \longrightarrow -15$	toe: 1.25	toe: $-7.36 \longrightarrow 15$	toe: 2.03
bump_2_h	14	$19 = 4 + 6 + 5 + 4$	toe: $0 \longrightarrow -15$	toe: 2.5	toe: $-7.36 \longrightarrow 15$	toe: 4.08
hole_1_h	24	$31 = 4 + 12 + 11 + 4$	toe: $7.36 \longrightarrow -15$	toe: 2.03	toe: $-15 \longrightarrow 15$	toe: 1.25
hole_2_h	13	$19 = 4 + 6 + 5 + 4$	toe: $5.4 \longrightarrow -15$	toe: 4.08	toe: $-15 \longrightarrow 15$	toe: 2.5
bump_1_comb	32	$31 = 4 + 12 + 11 + 4$	heel: $0 \longrightarrow -9$ toe: $0 \longrightarrow -15$	heel: 0.75 toe: 1.25	heel: $-9 \longrightarrow 9$ toe: $-7.36 \longrightarrow 15$	heel: 1.63 toe: 2.03
bump_2_comb	19	$19 = 4 + 6 + 5 + 4$	heel: $0 \longrightarrow -9$ toe: $0 \longrightarrow -15$	heel: 1.5 toe: 2.5	heel: $-9 \longrightarrow 9$ toe: $-7.36 \longrightarrow 15$	heel: 3.6 toe: 4.08
hole_1_comb	22	$31 = 4 + 12 + 11 + 4$	heel: $9 \longrightarrow -9$ toe: $7.36 \longrightarrow -15$	heel: 1.63 toe: 2.03	heel: $-15 \longrightarrow 9$ toe: $-15 \longrightarrow 15$	heel: 0.75 toe: 1.25
hole_2_comb	15	$19 = 4 + 6 + 5 + 4$	heel: $9 \longrightarrow -9$ toe: $5.4 \longrightarrow -15$	heel: 3.6 toe: 4.08	heel: $-15 \longrightarrow 9$ toe: $-15 \longrightarrow 15$	heel: 1.5 toe: 2.5

hole. This, however, is also due to the fact that the temporal informations taken individually already provided a high recognition rate.

5. EXPERIMENT 2

5.1. Description of the Conditions

In experiment 2, amplitude variations were involved while the time intervals both between footsteps and between heel and toe were kept fixed. The amplitudes patterns used were designed to simulate 14 different surface profiles. Specifically 2 flat, 6 bumps and 6 holes were designed. Such patterns involved three types of amplitude variations: amplitude variation of the toe, of the heel, and of both of them.

We designed our sounds in such a way that the amplitude decreased on ascending slopes and increased on descending slopes. In particular, the ranges of variations for heel and toe amplitudes were different. More in detail, the ranges of amplitude variation were chosen in order to reach the effect that when heel and toe variations were used together, on ascending slopes the heel amplitude was bigger than the toe one, while on descending slopes the heel amplitude was lower than the toe one. For this purpose the initial amplitude of the toe was normalized to the same amplitude of the heel (54 dB (A) for the wood and 58 dB (A) for the gravel), and an equal amplitude of heel and toe indicated the walk on a flat surface.

The characteristics of the 14 files used to drive the sound engine are illustrated in Table 4. In such table the suffixes _t, _h, and _comb indicate the type of amplitude variation used for each file (toe amplitude constant, heel amplitude constant, and the combinations of heel and toe variations respectively). During the variation of the amplitude of the toe, the amplitude of the heel remained constant, similarly during the variation of the amplitude of the heel the amplitude of the toe remained constant. The equations in the "Number of steps" column indicate how the steps where placed in reference to Fig. 2.

In order to model two different types of bumps and holes, for each category of surface modeling two slopes where chosen. The sound engine was set in order to synthesize footstep sounds on two different kinds of materials: wood and gravel. Participants were exposed to 28 trials, where the 14 surface profiles were presented twice in randomized order (each surface profile was presented with both wood and gravel).

5.2. Results Experiment 2

Tables 5 and 6 illustrate the results of experiment 2 for wood and gravel respectively. In both tables, the first column shows the different conditions as described in Table 4. The second, third and fourth columns illustrate the choices of the subjects (bump, hole or flat) for the different conditions they were exposed to. The fifth, sixth and seventh column report the average certainty expressed by the subjects after performing their choice; the fifth column reports the total certaintly in both correct and wrong answers, while the sixth and seventh column report the certainty in correct and uncorrect answers respectively. Finally, the last column reports the percentage of correct answers.

As the tables show, subjects could not successfully recognize bumps and holes using only the auditory cues described in the previous section. In fact, as can be seen in

Table 5 Results of experiment 2 for the wood surface. For a detailed description, see the text.

	Bump	Hole	Flat	Mean certainty			% Correct answers
				Total	Correct answers	Wrong answers	
flat_1		1	14	6.2	6.2857	5	93.33
flat_2		2	13	5.8667	6.0769	4.5	86.66
bump_1_t	3	8	4	4.8	4	5	20
bump_2_t	5	6	4	4.8	5.3	3.8	33.33
hole_1_t	4	4	7	4.7333	5	4	26.66
hole_2_t	9	5	1	4.6	3.4	5.2	33.33
bump_1_h	6	8	1	4.2667	3.25	4.6364	40
bump_2_h	3	11	1	4.6	4	4.75	20
hole_1_h	8	5	2	4.8	4	5.2	33.33
hole_2_h	7	4	4	4.0667	4.5	3.9091	26.66
bump_1_comb	3	12		5.1333	4.6667	5.25	20
bump_2_comb	7	7	1	4.7333	4.2857	5.125	46.66
hole_1_comb	7	4	4	4.2	4	4.2727	26.66
hole_2_comb	10	5		5	5	5	33.33

Table 6 Results of experiment 2 for the gravel surface. For a detailed description, see the text.

	Bump	Hole	Flat	Mean certainty			% Correct answers
				Total	Correct answers	Wrong answers	
flat_1	2	2	11	4.4	4.9091	3	73.33
flat_2	2		13	5.2667	5.5385	3.5	86.66
bump_1_t	7	5	3	4.4667	3.4286	5.375	46.66
bump_2_t	7	7	1	4.1333	3.4286	4.75	46.66
hole_1_t	5	6	4	4.3333	4.8333	4	40
hole_2_t	7	4	4	4.0667	2.5	4.6364	26.66
bump_1_h	5	6	4	4.1333	4.8	3.8	33.33
bump_2_h	5	7	3	4.3333	4	4.4545	33.33
hole_1_h	5	5	5	4.6667	3.6	5.2	33.33
hole_2_h	7	7	1	4.5333	4.2857	4.75	46.66
bump_1_comb	5	9	1	5.4	4.8333	5.7778	33.33
bump_2_comb	7	8		5.3333	4.8571	5.75	46.66
hole_1_comb	11	3	1	4.1333	3.6667	4.25	20
hole_2_comb	9	5	1	4.2	3.4	4.6	33.33

the last column of Tables 5 and 6, the percentage of correct answers is low for all conditions with exception of the flat surfaces. Indeed when bumps and holes were presented such percentages never reached the 50%.

An ANOVA was performed to examine whether significant differences were present in the recognition rate among the two surfaces and among the different conditions in the same surface. Overall, no significant differences were measured in the recognition rate among the two surfaces. Moreover, no significant differences were measured in the recognition rate for the different conditions in the same material. For example, no difference was measured in the recognition rate between the first and the second simulated hole.

Furthermore no significant difference was measured between the recognition rate obtained when changing the amplitude of the heel versus the one obtained when changing amplitude of the toe. Moreover, the combination of the two amplitude variations did not significantly enhance the recognition of a bump or a hole.

Observing columns 6 and 7, morever, it is possible to notice how subjects are not very certain when they express a correct answer or a wrong answer. In both surfaces, indeed, the mean certainty for both correct and wrong answers (without taking into account the flat surfaces) is not high (i.e., it is 4.3). More precisely such means for wood and gravel concerning the correct answers are 4.2 and 3.9 respectively, while for the wrong answers are 4.6 and 4.7 respectively. It is possible to notice a slight tendency in giving higher evaluations of the degree of certainty for wrong answers rather the correct ones, although all such differences are not statistically significative.

In addition an analysis performed on the results for each subject reveal that on average subjects were not consistent in their choices. Therefore is not possible to

Table 7 Results of experiment 3 for the wood surface with variations of amplitude and temporal distances. For a detailed description, see the text.

	Bump	Hole	Flat	Mean certainty			% Correct answers
				Total	Correct answers	Wrong answers	
flat_1			15	6	6	—	100
flat_2			15	6.1333	6.1333	—	100
bump_1_t	13		2	5.2667	5.4615	4	86.66
bump_2_t	13		2	4.4667	4.5385	4	86.66
hole_1_t	1	14		5.0667	5.2143	3	93.33
hole_2_t	3	12		4.6667	5	3.3333	80
bump_1_h	12	1	2	4.7333	4.8333	4.3333	80
bump_2_h	13	2		4.3333	4.6923	2	86.66
hole_1_h		13	2	5.2	5.3077	4.5	86.66
hole_2_h	1	13	1	4.6667	4.8462	3.5	86.66
bump_1_comb	13	1	1	4.6667	4.6923	4.5	86.66
bump_2_comb	14		1	4.8667	5.0714	2	93.33
hole_1_comb	4	11		4.8	5.2727	3.5	73.33
hole_2_comb	1	14		4.8667	5	3	93.33

deduct that they preferred the opposite mapping for the amplitude variation in rendering bumps and holes (i.e. increment of the amplitude on ascending slopes and decrement on descending slopes). This is also confirmed by the high percentages of flat choices when a bump or a hole was provided, as the fourth column of both the tables shows.

These results are very ambiguous, for this reason we can conclude that the proposed technique is not enough informative for rendering bumps or holes.

6. EXPERIMENT 3

6.1. Description of the Conditions

As mentioned in Section 3, in experiment 3 both the temporal intervals variations and the amplitude variations were involved. Among all the possible combinations of the stimuli involved in experiments 1 and 2, we chose to use as temporal patterns the combination of the temporal distance between footsteps and of the temporal distance between heel and toe, (i.e. the stimuli with the suffix _comb indicated in Table 1), and as amplitude variations the ones involved in experiment 2. Like in the experiments 1 and 2, in experiment 3 we simulated 14 different surface profiles (precisely 2 flat, 6 bumps and 6 holes were designed). Like in the previous experiments, the sound engine was set in order to synthesize footstep sounds on wood and gravel. Participants were exposed to 28 trials, where the 14 surface profiles were presented twice in randomized order (each surface profile was presented with both wood and gravel).

6.2. Results of Experiment 3

The results of experiment 3 for wood and gravel are shown in Tables 7 and 8 respectively. Such results are very similar to those of experiment 1. As the tables show, subjects could successfully recognize bumps and holes using the auditory cues described in the previous section. Indeed, the percentage of correct answers is high for all conditions, reaching also 100% of correct answers in two conditions (without considering the flat surfaces), and with a lowest score of 73% which was reached only in one condition in each table. Results shows that subjects were quite certain when they expressed a correct answer. In both surfaces, indeed, the mean certainty for correct answers is always above average. On the other hand, in situations where the answer was incorrect the degree of certainty is on average low.

An ANOVA was performed to examine whether significant differences were present in the recognition rate among the two surfaces and among the different conditions in the same surface. Such analysis brought to the same considerations made for the results of experiment 1 (see Section 4.2).

7. GENERAL DISCUSSION

From a comparison between the results of the three experiments it is possible to notice that the temporal variations are strongly dominant on the amplitude ones in providing the information about the simulated surface profiles. Indeed results of experiment 1 were very clear: subjects were able to recognize bumps and holes using the proposed auditory cues with high accuracy, both for gravel and wood. No significant differences were found in the recognition rate among the two surfaces nor among the different conditions in the same surface. No significant difference was present in the recognition rate for the two proposed temporal distances, and the combination of the two temporal information did not significantly enhance the recognition. This is especially due to the fact that the temporal informations taken individually already provided

Table 8 Results of experiment 3 for the gravel surface with variations of amplitude and temporal distances. For a detailed description, see the text.

	Bump	Hole	Flat	Mean certainty			% Correct answers
				Total	Correct answers	Wrong answers	
flat_1		1	14	5.4667	5.7143	2	93.33
flat_2	1		14	5.6667	5.9286	2	93.33
bump_1_t	13	1	1	4.6	4.4615	5.5	86.66
bump_2_t	13	2		4.8	5	3.5	86.66
hole_1_t	3	12		4.4667	4.5833	4	80
hole_2_t		15		5.2667	5.2667	—	100
bump_1_h	13	1	1	4.8667	5	4	86.66
bump_2_h	13	1	1	4.8	4.7692	5	86.66
hole_1_h	1	14		4.6667	4.7143	4	93.33
hole_2_h		15		5.4	5.4	—	100
bump_1_comb	12	1	2	4.4	4.75	3	80
bump_2_comb	11	2	2	4.6	5.3636	2.5	73.33
hole_1_comb	2	13		5.0667	5.3846	3	86.66
hole_2_comb	1	14		4.6667	4.8571	2	93.33

a high recognition rate. Concerning the degree of certainty of their answers, such evaluations were high for correct answers and low for the wrong ones, both for gravel and wood.

Conversely, in experiment 2 results were very ambiguous: subjects were not able to recognize bumps and holes using the proposed volume variations, and their degree of certainty in expressing an answer was not high, both for correct and wrong answers. Moreover, on average subjects were not consistent in their choices, as well as they slightly tended to answer with a flat choice when a bump or a hole was provided.

In addition results of experiment 3 show that the amplitude variations do not improve the recognition neither in terms of percentages of correct answers nor in terms of higher evaluations of degree of certainty. All the differences between experiments 1 and 3 concerning both the percentages and the evaluations of degree of certainty of correct answers, are not statistically significative. Therefore our hypothesis about an improvement of the recognition rate (and of higher evaluations of the degree of certainty) when using both the variations rather the single ones alone, was not confirmed.

As a consequence, it is possible to conclude that the proposed amplitude technique is not enough informative for rendering bumps or holes, while the variations of the temporal information alone are sufficient. This aspect can also be noticed in the results of experiment 2, where subjects frequently thought that the simulation was a flat surface, even if a bump or a hole was provided, thus confirming that the fixed temporal distances play an important role in the perception of walking over a flat surface.

This can also due to the fact that humans have a very high temporal resolution when exposed to auditory cues. On the other hand, distinguishing between variations in amplitude is not a straightforward task. In all the experiments the flat surfaces were always recognized with high accuracy and with high evaluations of degree of certainty. Furthermore in all the experiments no significative differences were found in the results concerning the two simulated materials, wood and gravel. Thus it is evident that in these types of tasks the use of an aggregate surface rather than a solid ones is not important, since it does not influence the recognition.

8. CONCLUSION AND FUTURE WORK

In this paper, we described a between subject experiment whose goal was to assess the role of temporal aspects and amplitude variations in recognition of some characteristics of footstep sounds, namely if a person is walking on a flat surface, a bump or a hole. The first experiment was performed only varying temporal parameters of footsteps, such as the distance between heel and toe and the distance between steps. The second experiment was run only varying amplitude parameters of footsteps, such as the amplitude of the heel and of the toe. In experiment 3 both the temporal and amplitude variations were combined. Results show that the temporal variations are strongly dominant on the amplitude ones in providing the information about the simulated surface profiles. In particular, results of experiment 1 show that participants were able to recognize with high accuracy bumps and holes provided by means of the proposed temporal variations. Conversely, results of experiment 2 are ambiguous, showing that participants were not successful in the recognition task using the proposed auditory cues. Moreover from a comparison between results of experiments 1 and 3 it is possible to notice that the variations of the amplitude did not improve the recognition neither in terms of percentages of correct answers nor in terms of higher evaluations of degree of certainty.

In consequence it is possible to conclude that the variations of the temporal information alone are sufficient for rendering bumps or holes, while the proposed amplitude technique is not enough informative.

Furthermore in all the experiments no significative differences were found in the results concerning the two simulated materials, and this is an indication that in these types of tasks the use of an aggregate surface rather than a solid ones does not influence the recognition.

The results presented in this paper have interesting applications in the field of navigations in virtual environments and computer games, where more realistic auditory feedback can enhance the simulated experience.

In future work we plan to render bumps and holes at auditory level interactively during the locomotion of a user.

ACKNOWLEDGMENTS

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.[‡]

REFERENCES

[1] M. Chion, C. Gorbman and W. Murch, "Audio-vision: Sound on screen," in *Columbia Univ Pr* (1994).
[2] X. Li, R. J. Logan and R. E. Pastore, "Perception of acoustic source characteristics: Walking sounds," *J. Acoust. Soc. Am.*, **90**, 3036 (1991).
[3] B. Giordano and R. Bresin, "Walking and playing: What is the origin of emotional expressiveness in music," *Proc. 9th Int. Conf. Music Perception and Cognition* (*ICMPC9*), *Bologna, Italy*, p. 149 (2006).
[4] R. Nordahl, S. Serafin and L. Turchet, "Sound synthesis and evaluation of interactive footsteps for virtual reality applications," *Proc. IEEE VR 2010*, pp. 148–153 (2010).
[5] A. DeWitt and R. Bresin, "Sound design for affective interaction," in *Affective Computing and Intelligent Interaction* (Springer, Berlin/Heidelberg, 2007), pp. 523–533.
[6] Y. Visell, F. Fontana, B. Giordano, R. Nordahl, S. Serafin and R. Bresin, "Sound design and perception in walking interactions," *Int. J. Hum.-Comput. Stud.*, **67**, 947–959 (2009).
[7] P. Cook, "Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds," *Proc. AES 22nd Int. Conf. Virtual, Synthetic, and Entertainment Audio*, pp. 73–78 (2002).
[8] F. Fontana and R. Bresin, "Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds," *Proc. Colloq. Musical Informatics*, pp. 109–114 (2003).
[9] A. J. Farnell, "Marching onwards: procedural synthetic footsteps for video games and animation," *Proc. Pure Data Convention* (2007).
[10] S. Serafin, L. Turchet and R. Nordahl, "Extraction of ground reaction forces for real-time synthesis of walking sounds," *Proc. Audiomostly Conf.*, pp. 99–105 (2009).
[11] N. E. Miner and T. P. Caudell, "Using wavelets to synthesize stochastic-based sounds for immersive virtual environments," *ACM Trans. Appl. Percept.*, **2**, 521–528, Association for Computing Machinery, Inc, One Astor Plaza, 1515 Broadway, New York, NY, 10036-5701, USA (2005).
[12] A. Lecuyer, J. M. Burkhardt and L. Etienne, "Feeling bumps and holes without a haptic interface: The perception of pseudo-haptic textures," *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pp. 239–246 (2004).
[13] M. Marchal, A. Lecuyer, G. Cirio, L. Bonnet and M. Emily, "Walking up and down in immersive virtual worlds: Novel interactive techniques based on visual feedback," *Proc. IEEE Symp. 3D User Interface*, pp. 19–26 (2010).
[14] S. Serafin, L. Turchet and R. Nordahl, "Do you hear a bump or a hole? An experiment on temporal aspects in footsteps recognition," *Proc. Digital Audio Effects Conf.*, pp. 169–173 (2010).
[15] L. Turchet, M. Marchal, A. Lecuyer, R. Nordahl and S. Serafin, "Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR," *Proc. ACM VRST* (2010).
[16] L. Turchet, S. Serafin, S. Dimitrov and R. Nordahl, "Physically based sound synthesis and control of footsteps sounds," *Proc. Digital Audio Effects Conf.*, 161–168 (2010).
[17] L. Turchet, R. Nordahl, A. Berrezag, S. Dimitrov, V. Hayward and S. Serafin, "Audio-haptic physically based simulation of walking sounds," *Proc. IEEE Int. Workshop Multimedia Signal Processing*, pp. 269–273 (2010).
[18] S. Serafin, L. Turchet, R. Nordahl, S. Dimitrov, A. Berrezag and V. Hayward, "Identification of virtual grounds using virtual reality haptic shoes and sound synthesis," *Proc. Eurohaptics Symp. Haptics and Audio-Visual Environments*, pp. 61–70 (2010).
[19] L. Turchet, R. Nordahl and S. Serafin, "Examining the role of context in the recognition of walking sounds," *Proc. Sound and Music Computing Conf.* (2010).

[‡]Natural Interactive Walking Project: www.niwproject.eu

Luca Turchet was born in Verona, Italy, where he received the Laurea degree (summa cum laude) in computer science from the University of Verona, in 2006. At the same time, he studied classical guitar and composition at the Music Conservatory E.F. Dall'Abaco of Verona receiving the Diploma degrees in 2007 and 2009 respectively (both summa cum laude). In 2006 he was visiting scholar at the Music Technology Group of Barcelona, Spain, and in 2008 at the Cork Institute of Technology — School of Music in Cork, Ireland. In 2009 he started the PhD at the Medialogy Department of the Aalborg University Copenhagen working on the Natural Interactive Walking Project. His main research interests are in interactive sound design, multimedia and multimodal systems, sound spatialization, and human-computer interaction.

Stefania Serafin is Professor in sound for multimodal environments at Aalborg University Copenhagen. She received a Ph.D. in Computer Based Music Theory and Acoustics from Stanford University in 2004, and a Master in Acoustics, Computer Science and Signal Processing Applied to Music from Ircam (Paris) in 1997. She has been visiting professor at the University of Virginia (2003), and visiting scholar at Stanford University (1999), Cambridge University (2002) and KTH Stockholm (2003). She is principal investigator for the EU funded project Natural Interactive Walking, and Danish delegate for the EU COST Action on Sonic Interaction Design. Her main research interests are sound models for interactive systems and multimodal interfaces, and sonic interaction design.