Do You Hear a Bump or a Hole?

An Experiment on Temporal Aspects in Footsteps Recognition

Serafin, Stefania; Turchet, Luca; Nordahl, Rolf

Published in:
Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

DO YOU HEAR A BUMP OR A HOLE? AN EXPERIMENT ON TEMPORAL ASPECTS IN THE RECOGNITION OF FOOTSTEPS SOUNDS

Stefania Serafin, Luca Turchet and Rolf Nordahl

Medialogy,
Aalborg University Copenhagen
Lautrupvang 15, 2750, Ballerup
`sts,tur,rn@media.aau.dk`

ABSTRACT

In this paper, we present a preliminary experiment whose goal is to assess the role of temporal aspects in sonically simulating the act of walking on a bump or a hole. In particular, we investigate whether the timing between heel and toe and the timing between footsteps affects the perception of walking on unflat surfaces. Results show that it is possible to simulate a bump or a hole by only using temporal information in the auditory modality.

1. INTRODUCTION

Sonic properties of footsteps have been extensively investigated in the auditory perception and sound design and synthesis community. From the perceptual point of view, previous research has shown that is it possible to recognize the gender of a human walker only by listening to recorded footsteps [10]. 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 [5]. Other studies demonstrates the possibility of recognizing simulated surfaces subjects are walking upon [13], as well as emotional intentions of the walker [2, 16].

To our knowledge, in the sound synthesis community all previous research on walking sounds has focused on the act of walking on flat surfaces [1, 4, 3, 14, 12]. Until recently, also the research on walking in virtual reality using visual cues has been focusing on flat surfaces. Few locomotion interfaces are able to render uneven ground, but they have the disadvantages of being costly and cumbersome [6, 7, 8]. Recently, research has shown that it is possible to simulate the act of walking on unflat surfaces only using visual cues [11]. These results are a development of previous research on pseudo-haptic simulation [9]. The main idea of the research described in [9] is to investigate whether it was possible to simulate a bump or a hole on a screen only by using visual feedback. This illusion is 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 is decelerating until reaching the top of the bump and then decelerating, while when simulating a hole the cursor first accelerates and then decelerates. Experiments show that subjects can successfully recognize a hole or a bump with this system [9].

Such research has recently been extended in [11], 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 the 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 show that such visualization techniques successfully simulate bumps and holes located in the ground.

In this paper, we are interested in exploring the possibility of implementing such pseudo-haptic feedback from the sonic point of view. While the 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 auditory cues. The research presented in this paper is part of the Natural Interactive Walking (NIW) FET-Open project[1], whose goal is to provide closed-loop interaction paradigms enabling the transfer of skills that have been previously learned in everyday tasks associated to walking. In the NIW project, several walking scenarios are simulated in a multimodal context, where especially audition and haptic play an important role. Until recently, the NIW project has been focusing on the simulation of the act of walking on flat surfaces. In this paper, we want to extend our simulations to account also for uneven surfaces such as bumps and holes.

2. SYNTHESIS OF FOOTSTEPS SOUNDS

In previous research, we proposed a sound synthesis engine able to simulate footsteps sounds on aggregate and solid surfaces both offline and in realtime [14, 13]. This engine is described in details in [15].

In all the different surfaces simulated, the system energy parameter is controlled by a ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. Such force is estimated from recordings on real footsteps sounds. [14].

For the purpose of this experiment, two types of surfaces, gravel and wood, were simulated, each with three characteristics: bumps, holes, and flat surfaces (see Figure 1). For each surface, the same footstep was used to create the different stimuli presented to the subjects, as described in the following section.

3. EXPERIMENT DESIGN

We conducted an experiment 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.

In a real environment, a person generally walks slower on ascending slopes, and faster on descending slopes. We transposed

[1] http://www.niwproject.eu/

this information in our experiment by modifying the time intervals both between footsteps and between the heel and toe information in each footstep.

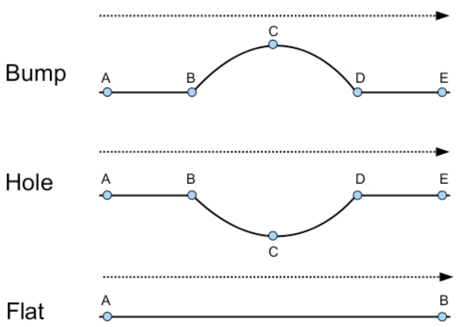

Figure 1: *The three types of surfaces modeled.*

3.1. Method

The sounds provided during the experiment consisted of footsteps sounds generated by the offline control of the sound engine. Some of the stimuli used in this experiment can be found at: `http://media.aau.dk/~sts/NIW/bumpsholes/`.

In order to simulate the wanted surface profiles, the input files for the engine were created placing at different temporal patterns a single footstep sound. The use of the same footstep was justified by the fact that we did not want other factors, such as changes in amplitude, to affect the results of the experiment. The footstep used to control the engine was extracted from a recording of a real walk on concrete floor using shoes with a solid soil. Figure 2 shows the waveform of the chosen footstep on top, and its corresponding GRF on the bottom. The sound was chosen among those available in the Hollywood Edge sound effects library.[2] 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 Figure 3).

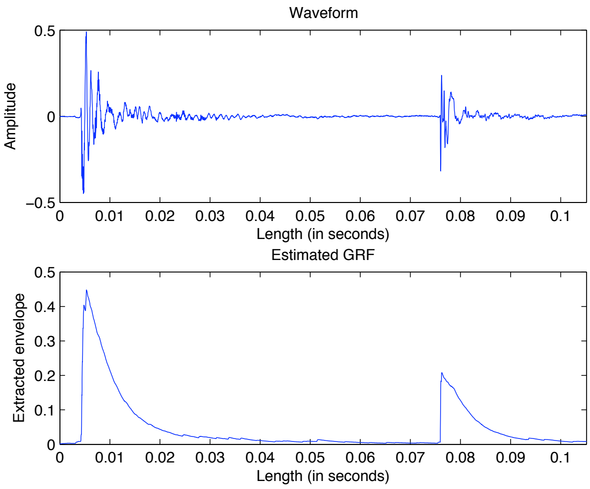

Figure 2: *Waveform of the used footstep on wood (top) and its relative extracted GRF (bottom).*

[2]www.hollywoodedge.com/

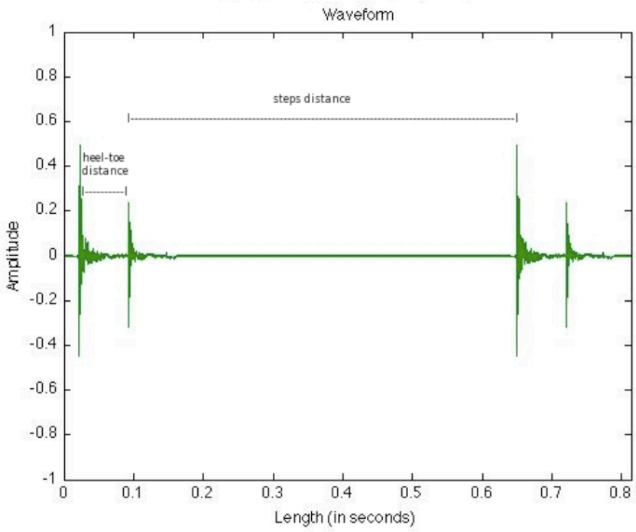

Figure 3: *Temporal distances between (named "steps distance" in the Figure), and within (named "heel-toe distance" in the Figure) footsteps.*

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 Figure 1. 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. Participants were exposed to 28 trials, where the 14 surface profiles were presented twice in randomized order. The sound engine was set in order to synthesize footsteps sounds on two different kinds of materials: wood and gravel. Each surface profile was presented with both wood and gravel.

The reason for choosing two materials was to assess whether the surface type affected the quality of the results. In this particular situation, a solid and an aggregate surface were chosen.

	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 ⟶ 1150	-	69 (fixed)
bump_2_step	16	19 = 4+6+5+4	100	550 ⟶ 1150	-	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 ⟶ 1150	30	0 ⟶ 360 (+ 69)
bump_2_comb	19	19 = 4+6+5+4	100	550 ⟶ 1150	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)

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

3.1.1. Participants

The experiment was performed by 15 participants, 11 men and 4 women, aged between 20 and 29 (mean=23.6,standard deviation=2.84). 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 minutes to complete the experiment.

3.1.2. Setup

All experiments were carried out in an acoustically isolated laboratory where the setups for the experiment was installed. It 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[3] 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.

3.1.3. Task

Subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface mentioned in section 3.1.2.

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 surfaces 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).

When moving to the next stimulus they could not change the answer to the previous stimuli.

[3]www.cycling74.com

4. RESULTS

The results of the experiments for wood and gravel are shown in Table 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, as described in the previous section; 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 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 certainly 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.

A t-test 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

	Bump	**Hole**	**Flat**	**Mean certainty Total**	**Mean certainty Correct answers**	**Mean certainty Wrong answers**	**% Correct 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 2: Results of the experiment for the wood surface.

	Bump	**Hole**	**Flat**	**Mean certainty Total**	**Mean certainty Correct answers**	**Mean certainty Wrong answers**	**% Correct 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

Table 3: Results of the experiment for the gravel surface.

did not significantly enhance the recognition of a bump or a hole. This, however, is also due to the fact that the temporal informations taken individually already provided a high recognition rate.

5. CONCLUSIONS

In this paper, we described an experiment whose goal is to assess the role of temporal aspects in recognition of some characteristics of footsteps sounds, namely if a person is walking on a flat surface, a bump or a hole. All experiments were performed only varying temporal parameters of footsteps, such as the distance between heel and toe and the distance between steps.

Other important aspects, such as amplitude and spectral information, were not modified. These parameters are object of future investigations. We are also planning to combine the results of our experiments with the results presented in [11], to assess how the combination of auditory and visual information affects the perception of walking on a hole or a bump.

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.

6. 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. [4]

7. REFERENCES

[1] P. Cook. Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds. *Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio*, pages 73–78, 2002.

[2] A. DeWitt and R. Bresin. Sound design for affective interaction. *Affective Computing and Intelligent Interaction*, pages 523–533, 2007.

[3] A.J. Farnell. Marching onwards: procedural synthetic footsteps for video games and animation. *Proceedings of the Pure Data Convention*, 2007.

[4] F. Fontana and R. Bresin. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. *Proc. Colloquium on Musical Informatics*, pages 109–114, 2003.

[5] B. Giordano and R. Bresin. Walking and playing: What is the origin of emotional expressiveness in music. In *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9), Bologna, Italy*, 2006.

[6] J.M. Hollerbach, D. Checcacci, H. Noma, Y. Yanagida, and N. Tetsutani. Simulating side slopes on locomotion interfaces using torso forces. In *Haptic Symposium*, pages 91–98. Citeseer, 2003.

[7] J.M. Hollerbach, R. Mills, D. Tristano, R.R. Christensen, W.B. Thompson, and Y. Xu. Torso force feedback realistically simulates slope on treadmill-style locomotion interfaces. *The International Journal of Robotics Research*, 20(12):939, 2001.

[8] H. Iwata, H. Yano, and F. Nakaizumi. Gait master: A versatile locomotion interface for uneven virtual terrain. *vr*, page 131, 2001.

[9] A. Lécuyer, J.M. Burkhardt, and L. Etienne. Feeling bumps and holes without a haptic interface: the perception of pseudo-haptic textures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 246. ACM, 2004.

[10] X. Li, R.J. Logan, and R.E. Pastore. Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America*, 90:3036, 1991.

[11] M. Marchal, A. Lecuyer, G. Cirio, L. Bonnet, and Emily M. Walking Up and Down in Immersive Virtual Worlds: Novel Interactive Techniques Based on Visual Feedback. In *Proceedings of IEEE Symposium on 3D User Interface*, 2010.

[12] N.E. Miner and T.P. Caudell. Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. *ACM Transactions on Applied Perception*, 2(4):521–528, 2005.

[13] S. Serafin R. Nordahl and L. Turchet. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proceedings of IEEE Virtual Reality*, 2010.

[14] S. Serafin, L. Turchet, and R. Nordahl. Extraction of ground reaction forces for real-time synthesis of walking sounds. In *Proc. Audiomostly*, 2009.

[15] L. Turchet, S. Serafin, S. Dimitrov, and R. Nordahl. Physically based sound synthesis and control of footsteps sounds. In *Proc. of Digital Audio Effects (DAFX) conference*, 2010.

[16] Y. Visell, F. Fontana, BL Giordano, R. Nordahl, S. Serafin, and R. Bresin. Sound design and perception in walking interactions. *International Journal of Human-Computer Studies*, 2009.

[4] www.niwproject.eu